MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 8: 381–410 (2024)
DOI: 10.3897/mbmg.8.127959

Research Article

Application of a dual-locus metabarcoding approach for a more comprehensive account of cattle dietary items in a semi-arid African savanna with special reference to forbs

Danielle Botha[1], Sandra Barnard[1], Morné Du Plessis[2], Mushal Allam[3,4], Kai Behn[5], Arshad Ismail[3,6,7], Anja Linstädter[8], Malesela Vincent Mokoka[9,10], Zamantungwa Thobeka Happiness Mnisi[11,12], Frances Siebert[1]

1 *Unit for Environmental Services and Management, North-West University, Potchefstroom, South Africa*
2 *Department of Genetics, University of the Free State, Bloemfontein, South Africa*
3 *Sequencing Core Facility, National Institute for Communicable Diseases, A Division of the National Health Laboratory Service, Johannesburg, South Africa*
4 *College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, United Arab Emirates*
5 *Institute of Crop Science and Resource Conservation, University of Bonn, Bonn, Germany*
6 *Department of Biochemistry and Microbiology, Faculty of Science, Engineering and Agriculture, University of Venda, Thohoyandou, South Africa*
7 *Institute for Water and Wastewater Technology, Durban University of Technology, Durban, South Africa*
8 *Biodiversity Research and Systematic Botany Group, Institute for Biochemistry and Biology, University of Potsdam, Maulbeerallee, Potsdam, Germany*
9 *Risk and Vulnerability Science Center, University of Limpopo, Sovenga, South Africa*
10 *Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa*
11 *Department of Biodiversity, University of Limpopo, Sovenga, South Africa*
12 *Department of Veterinary Tropical Diseases, University of Pretoria, Pretoria, South Africa*

Corresponding author: Danielle Botha (daniellebotha3@gmail.com)

OPEN ACCESS

Academic editor: Dirk Steinke
Received: 20 May 2024
Accepted: 23 October 2024
Published: 22 November 2024

Citation: Botha D, Barnard S, Du Plessis M, Allam M, Behn K, Ismail A, Linstädter A, Mokoka MV, Mnisi ZTH, Siebert F (2024) Application of a dual-locus metabarcoding approach for a more comprehensive account of cattle dietary items in a semi-arid African savanna with special reference to forbs. Metabarcoding and Metagenomics 8: e127959. https://doi.org/10.3897/mbmg.8.127959

Copyright: © Danielle Botha et al.
This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Abstract

Increasing livestock densities and more severe drought events challenge sustainable management in South Africa's semi-arid savannas. Effective mitigation strategies require accurate assessments of livestock foraging behaviour. By utilising high-throughput sequencing technology, this study evaluated the use of a dual-locus metabarcoding approach (*trn*L and *rbc*L) together with study-area-specific reference libraries, to analyse cattle diets in two bioregions of the eastern semi-arid South African savanna. Both markers demonstrated the ability to identify various plant families, but *trn*L exhibited a higher diversity in terms of family and genus identification at both sampling sites. Forbs, although comprising a diverse component of savanna plant communities, have relatively small above-ground biomass, but can still serve as crucial forage items, especially during dry periods. Our study underscores the significant role of forbs in cattle diets, demonstrating a shift in cattle foraging preferences from grass-based diets to higher inclusions of forbs and woody taxa during the drier season. Although grasses, such as *Setaria*, were still prevalent, forbs, belonging to the genera *Malvastrum*, *Asparagus*, *Pollichia* and *Ipomoea* were also important food items for cattle as well as woody taxa belonging to Fabaceae, Combretaceae, Ebenaceae, and Malvaceae with a selection of food items from trees and shrubs from genera *Albizia*, *Combretum*, *Euclea* and *Vachellia*. Furthermore, our study highlights the value of a dual-locus metabarcoding approach for understanding herbivorous diets. Using *trn*L and *rbc*L markers, with study-area-specific reference libraries, improves taxonomic resolution for accurately reconstructing cattle diets in semi-arid savannas. This study may improve biodiversity estimates and inform sustainable rangeland management strategies in semi-arid African savanna ecosystems.

Key words: African savanna, dietary analysis, forbs, semi-arid, *rbc*L, sustainable land-use, *trn*L

Introduction

African savannas cover almost half of the continent (Higgins et al. 1999; Gill et al. 2019) and semi-arid subtropical savannas are one of the most extensive vegetation types in South Africa (Shackleton et al. 2002). This vegetation type supports livestock production in large rural communities living on communal lands in South Africa (Shackleton et al. 2002; Monkwe et al. 2023). Communal cattle farming is particularly important to produce meat for domestic use, rituals, and to generate income (Monkwe et al. 2023). The Census of Commercial Agriculture report showed that in 2017 the total number of farming units involved in the commercial agriculture industry was 40 122, of which the largest proportion of farms was dedicated to livestock farming (13 639 or 33.9% of the total), followed by mixed farming (12 458 or 31.1%). Statistical data published by the Department of Agriculture Land Reform and Rural Development (2022) showed that in the growth period 2021/2022 South Africa had 12.2 million cattle.

Although semi-arid African savannas co-evolved together with large mammalian herbivores, the replacement of functionally diverse native African herbivore communities by single-species livestock has led to substantial changes in savanna vegetation structure and composition across the African continent (Hempson et al. 2017). One of the reasons is that livestock-based management practices usually aim at the improvement of grass biomass productivity (Gerrish 2004; Lemaire et al. 2011; Garcia et al. 2018). The traditional understanding of foraging ecology for rangeland management is based on the broad view of grass diets for grazers and tree-shrub forage for browsers, with limited recognition and understanding of the role of forbs in livestock and large herbivore communities foraging on grasslands (Craine et al. 2015). Studies that have examined bison diets showed that, under nutritional pressure, bison consumed fewer grasses (graminoids) and increased their intake of shrubs and forbs (Craine et al. 2015; Craine 2021). This shift in foraging behaviour allowed them to eat more nutritional plants with a higher protein content (Craine et al. 2015; Craine 2021). Forbs (i.e. non-graminoid flowering plants in the herbaceous layer with no or limited above-ground lignification) comprise the most diverse component of savanna plant communities (Trollope et al. 2014; Siebert and Scogings 2015; Siebert et al. 2020) although their contribution to above-ground biomass usually remains minor compared to grasses (Linstädter et al. 2014). Yet, forbs form part of the diet of several native savanna herbivores (Pansu et al. 2022) and may be seasonally important forage items for which browsing and mixed-feeding wildlife may compete with grazing livestock (Odadi et al. 2013), especially during dry periods (Odadi et al. 2007). Understanding the dietary preferences of grazing animals is needed to improve their nutrition, health, and welfare (Villalba et al. 2015; Mysterud and Austrheim 2016; Distel et al. 2020). These insights are also crucial to preserving the health, integrity and resilience of the ecosystems they inhabit (Villalba et al. 2015; Rivero et al. 2021) as well as guiding the proactive management of toxic plants (Scasta et al. 2020).

Environmental factors affecting current forage availability and seasonal variation in a grazer's diet interact with individual animal traits in influencing diet composition (Mysterud and Austrheim 2016). With the projected increase in the frequency and intensity of drought events (Ayugi et al. 2022; Chivangulula et al. 2023) and the related pressure on food security, an improved understanding of alternative indigenous forage items for livestock is key to future ecosystem management, particularly in dryland rangelands (Kleppel 2020; Mudzengi et al. 2020). Recent technological developments, based on high-throughput sequencing, have made new perspectives on dietary analysis and biodiversity assessments possible (Katikati and Fourie 2019). The estimation of animal diets by faecal DNA metabarcoding has been recognised as a non-intrusive, precise, time- and cost-efficient method. The high taxonomic resolution displayed by this technique (Nichols et al. 2016) has enabled the detection of numerous food items that are challenging to identify by traditional methods (Soininen et al. 2009; Bhattacharyya et al. 2018), allowing for the exact mechanisms of niche partitioning in numerous sympatric species to be revealed (Kartzinel et al. 2015; Pansu et al. 2018; Ter Schure et al. 2021; Aylward et al. 2022) and showed several crucial elements of wildlife management such as food selection of endangered species (Gebremedhin et al. 2016; Iwanowicz et al. 2016; Goldberg et al. 2020). Faecal DNA analysis can also inform farm management on applying adaptive grazing strategies and the seasonality of grazing to improve livestock gain responses (Jorns et al. 2023).

However, selecting which genomic region, or DNA barcode, to use for species identification in a metabarcoding study is crucial to retrieving and identifying target taxa from an environmental sample. These barcodes are chosen for their sequence recovery, annotation, and sequence resolution among taxa (Hajibabaei et al. 2016). While the mitochondrial cytochrome c oxidase subunit 1 (COI) gene represents an exemplary barcode for the identification of animals (Coissac et al. 2016), plants possess no such single candidate marker exhibiting the same characteristics (Yu et al. 2011). Furthermore, it was emphasised that a barcode standard must be augmented in some studies, such as dietary analysis due to the required ability of the barcoding marker to amplify degraded and low copy number DNA from faecal samples (De Groot et al. 2011; Hollingsworth et al. 2011). In the case of herbivore diet analysis, relatively few studies include the use of the core plant barcodes, *rbc*L and *mat*K, recommended by the CBOL Plant Working Group (García-Robledo et al. 2013; Kartzinel et al. 2015; Mallott et al. 2018). The *rbc*L and *trn*L(UAA) barcodes have been proposed as two single-locus gene regions for the analysis of herbivorous- and omnivorous animal diets via DNA metabarcoding (Taberlet et al. 2007; Valentini et al. 2009; Kartzinel et al. 2015; Kress et al. 2015; Mallott et al. 2018), and will constitute the core barcoding regions in this study. Advantageous characteristics of one of the barcoding loci recommended by CBOL, *rbc*L, include the universality of its primers (De Groot et al. 2011; Yang et al. 2014), the conservative coding nature of the region that can be used to reconstruct deep evolutionary relationships, high recoverability, and specimen identification up to family level (Taberlet et al. 2007; Ferri et al. 2015; Lee et al. 2018). *Trn*L is also a chloroplast gene and yields small DNA fragments with the whole intron (254–767 bp) representing the forward *c* and reverse *d* regions which are highly conserved among land

plants (Taberlet et al. 1991; Taberlet et al. 2007). The *rbc*L and *trn*L barcodes are also well represented on public reference databases with 280 050 *rbc*L and 325 408 *trn*L plant sequences on the non-redundant (nt) nucleotide database of the National Centre for Biotechnology Information (NCBI: Johnson et al. 2008) as of the 20th of March 2024. Although species-level identifications for plants constituting the diets of cattle at Welverdiend and Syferkuil would be optimal, it was decided to evaluate the diets to genus level for the following reasons: i) neither *rbc*L nor *trn*L has particularly high rates of species-level discrimination (Taberlet et al. 2007; Valentini et al. 2009; Hollingsworth et al. 2011); ii) operational taxonomic units (OTUs), such as those reported in dietary studies are difficult to determine at the species level, as previous research has shown (De Barba et al. 2014; Buglione et al. 2018; Scasta et al. 2019); and iii) it can be difficult to obtain an amplicon length long enough to reliably identify samples down to the species level (Deagle et al. 2010). This problem is mostly brought about by the degradation of food items in faecal samples (Deagle et al. 2006). The use of *rbc*L and *trn*L has shown great potential to identify plants to genus level in dietary studies (Yoccoz et al. 2012; Mallott et al. 2018; Matiz-Ceron et al. 2022; Thongjued et al. 2024).

This study aims to evaluate the use of a dual-barcode approach together with study-area-specific reference libraries (Botha et al. 2023) to determine and quantify the dietary composition of cattle foraging in two bioregions of the eastern semi-arid South African savanna, with special attention to forb species. By applying this approach to cattle faecal samples, key dietary items in terms of plant families and genera, and the relative importance of forbs as forage to these animals during the drought season were determined.

Materials and methods

Study area

The study sites that were used for the estimation of dietary profiles of domesticated cattle were located within the eastern semi-arid savanna in the Limpopo Province of South Africa (SA) across two bioregions, namely the Central Bushveld- and the Lowveld Bioregion (Mucina and Rutherford 2006). Welverdiend (24°34'15"S, 31°20'42"E) is a rural village located close to Bushbuckridge in the SA Lowveld which supports a communal rangeland along the periphery of the village. This area is predominantly populated by broad-leaved woody species, like *Senegalia nigrescens* (Oliv.) PJH Hurter, *Sclerocarya birrea* subsp. *caffra* (Sond.) Kokwaro, *Albizia harveyi* E Fourn., and various species from the Combretaceae family (Mucina and Rutherford 2006). The grass layer is mostly composed of species such as *Digitaria eriantha* subsp. *eriantha* (Steud.) Maire, *Eragrostis rigidior* Pilg., *Melinis repens* (Willd.) Zizka, *Schmidtia pappophoroides* Steud. ex JA Schmidt, *Urochloa mosambicensis* (Hack.) Dandy, and *Aristida congesta* Roem. & Schult. Additionally, the area boasts the forb species from genera such as *Commelina*, *Hibiscus*, *Sida*, and various others from the Malvaceae, Fabaceae, Convolvulaceae, and Acanthaceae families (Mucina and Rutherford 2006). Syferkuil (23°50'36.86"S, 29°40'54.99"E) is the experimental farm of the University of Limpopo where rotational grazing is being practised (Scheiter et al. 2024). The vegetation of the Syferkuil study area has been described as

an open thorn bush savanna, specifically belonging to the Pietersburg Plateau False Grassveld type (Acocks 1988). In this ecosystem, the herbaceous layer is mainly composed of perennial grass species such as *Themeda triandra* Forssk., *Digitaria eriantha* Steud., *Schmidtia pappophoroides* Steud. ex J.A. Schmidt, and various species of *Eragrostis*. Additionally, the woody component is predominantly represented by *Vachellia tortilis* (Forssk.) Galasso & Banfi (Low and Rebelo 1998). The climate of both sites is semi-arid with a mean annual rainfall of 350–500 mm, most of which occurs in the summer months (Shackleton 2000; Matsika et al. 2013; Phefadu and Kutu 2016; Munjonji et al. 2020). Comprehensive plant species lists were compiled as described by Botha et al. (2023). The *rbc*L and a *trn*L DNA barcode sequence reference libraries of these plants were developed and evaluated by Botha et al. (2023).

Sample collection, DNA isolation, amplification, and sequencing

DNA was extracted from the faeces of 24 individual cattle from Welverdiend and 24 individual cattle from the Syferkuil Experimental farm during the end of the dry season, specifically directly before the first rains of the new season when forage supply from the grass layer is particularly scarce (Linstädter et al. 2013). DNA was also extracted in the same manner from a blank control made up of molecular water, labelled as sample C2. The faecal samples from Welverdiend were collected in October 2018 and those from Syferkuil in October 2019. Great care was taken to ensure that these faecal samples were representative of individuals either by visual observation or by allowing a sufficient gap in time and space in the collection efforts. Approximately 10 g of fresh faecal material was collected after defecation per individual and ~200 mg of this was used for subsequent DNA isolation using the Qiagen DNeasy mini–Plant DNA Extraction kit according to the manufacturer's protocol. Faecal samples were collected without interacting with or disturbing animals.

The *rbcL* and *trn*L(UAA) intron chloroplast gene regions from 48 faecal samples as well as the extraction blank control, sample C2, were amplified using the barcode's metagenomic sequencing library protocol (Illumina). Two PCR reactions were completed for *rbc*L*a* and *trn*L(UAA) library preparations and incorporated the Illumina overhang adapters onto the primers: *rbc*L*a*-F (Levin et al. 2003), *rbc*L*a*-R (Kress and Erickson 2007), *trn*L*c* and *trn*L*d* (Taberlet et al. 1991), which are further described in Suppl. material 1. Each PCR reaction had a final volume of ~25 µl which consisted of 12.5 ng DNA template, 5 µl (1 µM) of the forward and reverse primer, respectively, 12.5 µl 2X Kapa HiFi HotStart ReadyMix (KAPA Biosystems Woburn, MA), and 2.5 µl PCR-grade water.

The PCR amplification of these reactions, as well as negative control reaction without any template DNA, were carried out simultaneously as follows: heated lid 110 °C, 95 °C for 30 seconds, 25 cycles of 95 °C for 30 seconds, 55 °C for 30 seconds, 72 °C for 30 seconds, then 72 °C for 5 minutes and held at 4 °C. Gel electrophoresis (1X TAE buffer, 1.5% agarose, 100 V) was used to visualise the PCR results and successful PCR products were purified using AMPure XP magnetic bead-based purification (Beckman Coulter, IN). Multiplexing was made possible by the Illumina Nextera XT Index kit (Illumina, San Diego, CA) using two 8-base indices. Two indexing primers per sample were utilised and PCR reactions with the purified DNA template (5 µl DNA template, 5 µl × 2 index

primer, 25 µl 2x KAPA HiFi HotStart ReadyMix and 10 µl PCR-grade water) were performed on a Bio-Rad C1000 Thermal Cycler (Bio-Rad, Hercules, CA). The cycling conditions consisted of an initial denaturation cycle of 95 °C for 3 minutes, followed by eight cycles of denaturation (95 °C for 30 seconds), annealing (55 °C for 30 seconds), and extension (72 °C for 30 seconds). The cycle was completed with a final extension cycle of 72 °C for 5 minutes. The indexed libraries were purified with Ampure XP beads before library pooling, and their quantities were determined using the Qubit dsDNA HS Assay Kit from Thermo Fisher Scientific (Waltham, MA). Prior to sequencing, the quality of the purified amplicons was evaluated using the Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA). Following instructions from Illumina, the sample pool (2 nM) was denatured with 0.2 N NaOH, diluted to 7 pM, and mixed with 10% (v/v) denatured 20 pM PhiX. Samples were sequenced on the Illumina MiSeq platform (2 × 300 bp) by the NICD Sequencing Core Facility service, in Pretoria, South Africa.

Bioinformatics analysis

Cutadapt (v. 4.1: Martin 2011) was used to remove PCR primers from the *rbcL* and *trnL* amplicons and additional quality filtering, trimming, sample inference, dereplication and the removal of chimeric sequences with the "dada2" R package (v.1.24: Callahan et al. 2016) on forward reads. The default parameters of DADA2 included the removal of sequences containing ambiguous bases, sequences that were truncated at base pair positions at the first instance of quality scores less than 20, and sequences that were shorter than 50 bp. Default parameters were therefore used in most cases, except that reads were not merged. The reason for this is that a minimum of 20 base pairs are required to overlap to create a consensus sequence and it would be unsuccessful for larger amplicons expected for the *trnL* intron. This choice is not unusual, especially when large amplicons or amplicons with highly variable lengths or low-quality reverse reads are present as seen for both the *rbc*L and *trn*L reads (Pauvert et al. 2019; Tan et al. 2019). Data handling hereafter will therefore only be applied to unmerged, forward reads. The amplicon sequence variants (ASVs) left after chimaera removal and resulting from sample inference were subject to additional filtering and curation.

Frequency-based filtering was applied to both the *rbc*L and *trn*L datasets, which aided in removing ASVs only present in a single sample and an additional abundance filter, identified as necessary when tested with a simulated mock community (Botha 2023 and Suppl. material 2), was applied to the *rbc*L datasets which removes ASVs representing 0.005% of the total dataset (Bokulich et al. 2013). The removal of these low-frequency features, which may represent contamination, PCR, or sequencing errors, results in decreased false positives and a truer reflection of animal foraging behaviour. The filtered ASV tables for *rbc*L and *trn*L were then subject to post-clustering curation, or post-denoising in the case of DADA2, by the LULU algorithm. The R package “lulu” (v.0.1: Frøslev et al. 2017) identifies and merges often erroneous ‘daughter’ OTUs with consistently abundant co-occurring sequences, or ‘parent’ OTUs (Frøslev et al. 2017). Curated ASVs remaining in the feature tables for *rbc*L and *trn*L were assigned to the taxonomy contained in a custom, study-area-specific reference libraries (Botha et al. 2023) using the “blast*n*” command under the “BLAST+” tool to

only keep one identification per ASV that has similarity threshold of 80% and an e-value of 0.00001. The *rbc*L reference library consisted of 1 238 sequences representing 318 genera and 562 species. The *trn*L reference library consisted of 921 sequences representing 270 genera and 461 species.

As part of an additional quality control step, the unassigned ASVs resulting from taxonomic assignment by the compiled reference libraries by Botha et al. (2023) were "BLAST searched" against the contents of NCBI's non-redundant nucleotide (nt) database on the 13th of September 2022. This was done to establish (i) whether the unassigned reads had non-plant origins, and (ii) whether the unassigned ASVs are attributed to poor taxonomic coverage of the reference libraries. Reads that were assigned to species other than plants, with a percentage identity of 80%, query coverage of 80% and e-values smaller than 0.001 were removed from the feature tables for each dataset since they were not a reflection of the inability of the reference libraries to identify the composition of herbivores foraging in an eastern semi-arid South African savanna and cannot truly be classified as "unassigned".

To determine the sum of branch lengths (SBL) of the *trn*L and *rbc*L datasets, the ASVs were aligned in the Multiple alignment program for amino acid or nucleotide sequences (MAFFT, v.7: Katoh and Standley 2013) and then used to reconstruct neighbour-joining (NJ) phylogenetic trees for the *trn*L and *rbc*L datasets, as implemented by the package "ape" (v.5.6.2: Paradis and Schliep 2019) in R (v.4.3.1: R Core Team 2023) with bootstrap testing of 1 000 replicates using the Jukes-Cantor model. SBL was calculated for both the assigned and the unassigned reads after the removal of non-plant sequences, frequency and abundance-based filtering and the application of the LULU algorithm for the *rbc*L and *trn*L amplicon per bioregion.

Statistical analyses

Alpha-diversity was determined using the Shannon index and was calculated based on the abundance matrices obtained for each primer dataset for both assigned and unassigned reads sampled from each bioregion using the estimate-richness function of the "phyloseq" (v.1.48: McMurdie and Holmes 2013) package. All analyses were performed using ASVs that were taxonomically identified by the reference libraries created by Botha et al. (2023). The Shannon index provided insights into the species richness detected by each marker within each treatment group. The Kruskal-Wallis test was used to compare multiple independent groups within the filtered datasets. It was performed using R software to evaluate the statistical difference in plant species richness and diversity between the different markers and bioregions. Plots were created using the package "ggplot2" (v.3.3.6: Wickham 2016). The p-values less than 0.05 were considered significant.

Dietary richness evaluation

Taxonomic assignments of the *rbc*L and *trn*L ASVs retrieved from the Syferkuil and Welverdiend cattle faecal samples were visualised via read abundance, frequency of occurrence (FOO), also known as absolute data, as well as the relative read abundance (RRA) of ASVs assigned to family and genera per

sampled individual. Each taxon was grouped according to plant life forms, i.e., woody taxa (trees/shrubs), woody taxa/forbs (shrubs/forbs), forbs, and graminoids (grasses). All statistical analyses were carried out in R (v.4.3.1: R Core Team 2023). Absolute read abundances were calculated and visualised as bar plots with the "phyloseq" package to show the number of detected ASVs for each marker per study site. These bar plots were employed to summarise and visualise the number and proportion of assigned and unassigned ASVs (NA reads). Distinct approaches regarding treating unassigned reads were used for FOO and RRA calculations and visualisations to ensure a robust analysis and accurate portrayal of taxonomic proportions across samples. NA reads were omitted for FOO calculations to focus on the occurrence of taxa that could be assigned to the respective *rbc*L and *trn*L datasets. The FOO percentages of families and genera were calculated by summing all present values per specified family or genus (1 = present; 0 = absent) for samples in the dataset, dividing it by the number of samples and multiplying this total by 100. FOO used bar graphs to visualise the data. In contrast, RRA calculations and visualisations retained NA reads to accommodate for accurate interpretations of taxa proportions. NA reads are prevalent in amplicon sequencing datasets due to the absence of taxa in reference databases or inadequate and differential amplification of certain taxa by barcoding primers. The omission of NA reads from RRA visualisations (100% stacked bar plots) would lead to discrepancies in the interpretation of taxa proportions, particularly when comparing samples with different degrees of NA read removal. Furthermore, each sample was assigned to a dominant life form category (Graminoids, Forbs, Forbs/Woody taxa and Woody taxa) to contextualise RRA within a dietary analysis framework. In cases where NA reads were dominant, the second most abundant life form assigned to families or genera was used for categorisation. Samples were then grouped into facets and arranged according to the descending order of abundance of these dominant life forms. RRA percentages were reported based on taxa proportions of assigned reads constituting each sample and were calculated by dividing the count of each unique ASV belonging to a specific genus or family in a sample by the sum of all assigned ASVs in that sample. RRA and FOO bar graphs were created with the "ggplot2" package.

Results

Metabarcoding sequence summary

The Welverdiend *rbc*L dataset comprised of 10 209 509 sequences and 4 989 550 ASVs, of which 494 were left after filtering and post-clustering curation. For the *trn*L dataset, 3 894 691 sequences resulted in 1 852 442 ASVs of which 1 217 were left after filtering and post-clustering curation. The Syferkuil *rbc*L dataset consisted of 1 858 096 sequences and resulted in the creation of 1 444 146 ASVs of which 1 403 were left after filtering and curation. Of the 1 985 791 sequences that were input for the Syferkuil *trn*L dataset, 1 234 380 ASVs were created and 3 320 were left after filtering and trimming. ASVs that were assigned to non-plant references and were subsequently removed from further analysis included 17 from the Welverdiend *trn*L dataset, 51 from the Syferkuil *rbc*L dataset and 18 from the Syferkuil *trn*L dataset. Furthermore, the blast

results indicated that ASVs classified as unassigned by the compiled reference libraries could also not be matched with any plant species in NCBI's non-redundant database within the specified parameters.

Comparison of phylogenetic distances and alpha diversity metrics across foraged communities in the Welverdiend and Syferkuil sites

To compare the phylogenetic distances of the species detected with the *trn*L and *rbc*L at each of the foraged communities, Welverdiend and Syferkuil, NJ trees from each treatment for the *trn*L and the *rbc*L ASVs were compared. The phylogenetic distances or SBL for the *trn*L ASVs were higher than the *rbc*L ASVs at both sites. Specifically, at Welverdiend, the SBL for *trn*L (295.51) exceeded that for *rbc*L (137.58), while at Syferkuil, *trn*L (934.78) also exceeded that of *rbc*L (358.77). The SBL for both markers was higher at Syferkuil than at Welverdiend. The alpha diversity metrics (Shannon index) revealed that the highest diversity and richness of consumed plant species (both assigned and unassigned) was detected by the *trn*L marker at the Syferkuil sampling site (Fig. 1). This site also showed the highest diversity for the *rbc*L marker. Plant species richness differed significantly between bioregions and markers (Kruskal-Wallis test, $p < 0.05$).

Dietary analysis of cattle foraging at Welverdiend and Syferkuil

Total read abundance

For Welverdiend, the total abundance of unassigned reads was more prevalent in the *trn*L dataset than in the *rbc*L dataset (Suppl. material 3), although *trn*L exhibited a higher diversity in family identification. At Welverdiend both markers were able to identify the families Fabaceae, Anacardiaceae, Ebenaceae, Malvaceae, and Poaceae with *rbc*L exclusively identifying Combretaceae, Acanthaceae, and Annonaceae, and *trn*L identifying Apocynaceae, Asparagaceae, Asteraceae, Capparaceae, Convolvulaceae, Cyperaceae, Ehretiaceae, Euphorbiaceae, Moraceae, and Rubiaceae. All eight families identified by *rbc*L were identified to genus level with Fabaceae being represented by *Philenoptera*, *Albizia*, and *Dichrostachys*, Anacardiaceae by *Rhus*, Combretaceae by *Combretum*, Ebenaceae by *Euclea* and *Diospyros*, Annonaceaea by *Hexalobus*, Poaceae by *Enteropogon* and *Trachypogon* and Malvaceae by *Malvastrum*. With *trn*L, only seven families could be identified to genus level with Fabaceae being represented by *Senegalia* and *Vachellia*, Ebenaceae by *Euclea*, Poaceae by *Eragrostis* and *Setaria*, Asparagaceae by *Asparagus*, Convolvulaceae by *Ipomoea*, Ehretiaceae by *Ehretia* and Moraceae by *Ficus*.

The total abundance of unassigned ASVs for the Syferkuil *rbc*L dataset was much more prevalent than those that could be linked to a taxonomy and the dataset only consisted of 24 samples since the control sample was lost during filtering. The Syferkuil *trn*L dataset was also represented by only 24 samples since all reads from sample 43 were lost during filtering and this dataset also included the highest abundances of unassigned reads relative to the other three datasets. In the case of the Syferkuil dataset (Suppl. material 4), *rbc*L was able to identify more plant families than *trn*L. In the Syferkuil dataset, *rbc*L and *trn*L commonly identified

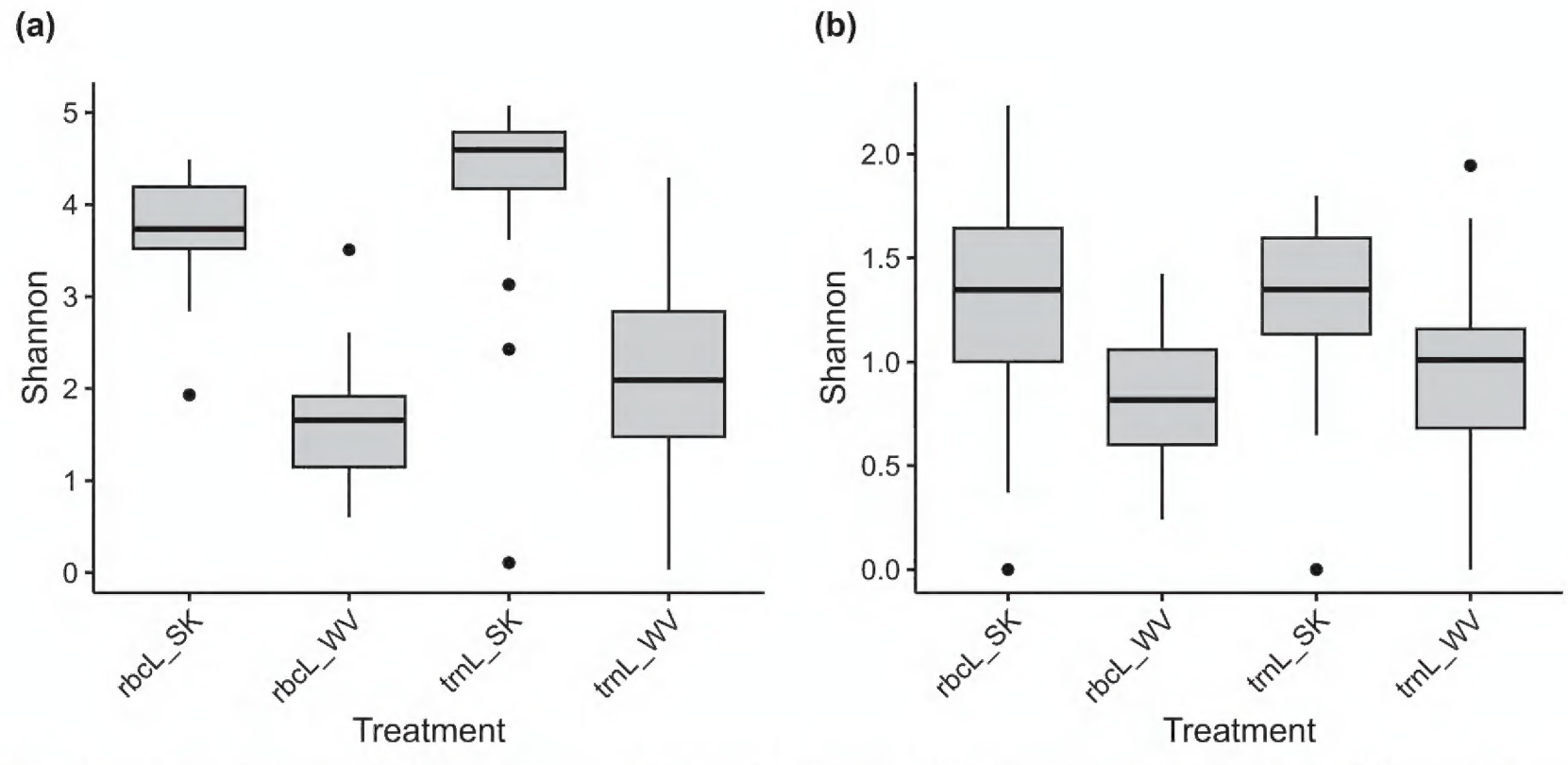

Figure 1. Boxplots representing Shannon Diversity Index for unassigned (**a**), and assigned ASVs (**b**), calculated per barcode (*rbc*L and *trn*L) for Welverdiend (WV) and Syferkuil (SK).

Caryophyllaceae, Ebenaceae, Fabaceae, Poaceae, and Solanaceae. *Rbc*L exclusively identified Apocynaceae, Geraniaceae, Phyllanthaceae, and Rhamnaceae while *trn*L could identify Cleomaceae, Euphorbiaceae, and Lamiaceae. All nine families identified by the *rbc*L marker could also be identified to genus level as seen with the Welverdiend dataset, with Caryophyllaceae represented by *Pollichia*, Apocynaceae by *Adenium*, Ebenaceae by *Diospyros*, Fabaceae by *Vachellia* and *Dichrostachys*, Geraniaceae by *Monsonia*, Phyllanthaceae by *Flueggea*, Poaceae by *Setaria*, *Cenchrus*, and *Tricholaena*, Rhamnaceae by *Berchemia* and *Ziziphus*, and Solanaceae by *Solanum*. Like in the Welverdiend dataset, *trn*L could not identify all eight families to genus level with Caryophyllaceae represented by *Pollichia*, Ebenaceae by *Euclea*, Fabaceae by *Indigofera*, *Senegalia*, *Senna* and *Vachellia*, Poaceae by *Setaria*, Solanaceae by *Solanum*, Cleomaceae by *Cleome* and Lamiaceae by *Leucas*.

The use of *rbc*L and *trn*L as a dual-locus barcoding approach revealed complementary capabilities in the reconstruction of cattle dietary profiles. Both markers could identify a range of plant families, with *trn*L exhibiting higher family diversity compared to *rbc*L in both Welverdiend and Syferkuil study areas. While *trn*L could uniquely identify families like Apocynaceae, Asteraceae, and Euphorbiaceae, *rbc*L could exclusively identify families such as Combretaceae and Annonaceae. Genus level identifications varied between the markers, with *rbc*L consistently identifying genera across all families whereas *trn*L had fewer representative genera of common families identified or no genera representative of unique families identified. Table 1 contains a comprehensive summary of plant families that could be recovered with the *rbc*L and *trn*L barcoding markers that make up the diets of cattle foraging at both Welverdiend and Syferkuil.

Percentage FOO and relative read abundances

The FOO analysis of both the *rbcL* and *trn*L datasets from Welverdiend (Fig. 2) showcased the diverse utilisation of various life forms within this semi-arid ecosystem at the end of the dry season, revealing discernible patterns of variations in the FOO among these life forms. The FOO% calculations for the Welverdiend

Table 1. A summary of the ability of the *rbc*L and *trn*L barcodes to recover plant life forms from cattle diets on family and genus levels.

Family	Genus	*rbc*L	*trn*L	Trees/ Shrubs	Forbs	Graminoids	Study area	
							WV	SK
Acanthaceae		✓			✓		✓	
	Thunbergia	✓			✓		✓	
Anacardiaceae		✓	✓	✓			✓	
	Rhus	✓		✓			✓	
Annonaceae		✓		✓			✓	
	Hexalobus	✓		✓			✓	
Apocynaceae		✓	✓		✓		✓	✓
	Adenium		✓		✓			✓
Asparagaceae			✓		✓		✓	
	Asparagus		✓		✓		✓	
Asteraceae			✓		✓		✓	
Capparaceae			✓	✓	✓		✓	
Caryophyllaceae		✓	✓		✓			✓
	Pollichia	✓	✓		✓			✓
Cleomaceae			✓		✓			✓
	Cleome		✓		✓			✓
Combretaceae		✓		✓			✓	
	Combretum	✓		✓			✓	
Convolvulaceae			✓		✓		✓	
	Ipomoea		✓		✓		✓	
Cyperaceae			✓			✓	✓	
Ebenaceae		✓	✓	✓			✓	✓
	Diospyros	✓		✓			✓	✓
	Euclea	✓	✓	✓			✓	✓
Ehretiaceae			✓	✓			✓	
	Ehretia		✓	✓			✓	
Euphorbiaceae			✓	✓	✓		✓	✓
Fabaceae		✓	✓	✓	✓		✓	✓
	Dichrostachys	✓		✓			✓	✓
	Vachellia	✓	✓	✓			✓	✓
	Albizia	✓		✓			✓	
	Philenoptera	✓		✓			✓	
	Indigofera		✓		✓			✓
	Senegalia		✓	✓			✓	✓
	Senna		✓		✓			✓
Geraniaceae		✓			✓			✓
	Monsonia	✓			✓			✓
Lamiaceae			✓		✓			✓
	Leucas		✓		✓			✓
Malvaceae		✓	✓	✓	✓		✓	
	Malvastrum	✓	✓		✓		✓	

Family	Genus	*rbc*L	*trn*L	Trees/ Shrubs	Forbs	Graminoids	Study area	
							WV	SK
Moraceae			✓	✓			✓	
	Ficus		✓	✓			✓	
Phyllanthaceae			✓	✓				✓
	Flueggea	✓		✓				✓
Poaceae		✓	✓			✓	✓	✓
	Cenchrus		✓			✓		✓
	Setaria	✓	✓			✓	✓	✓
	Tricholaena	✓				✓		✓
	Enteropogon	✓				✓	✓	
	Trachypogon	✓				✓	✓	
	Eragrostis		✓			✓	✓	
Rhamnaceae		✓		✓				✓
	Berchemia	✓		✓				✓
	Ziziphus	✓		✓				✓
Rubiaceae			✓		✓		✓	
Solanaceae		✓	✓		✓			✓
	Solanum	✓	✓		✓			✓

*rbc*L dataset showed a consistent presence of woody taxa across the majority of samples. *Albizia* and *Combretum* were predominant (FOO = 96%), whereas forbs, *Malvastrum* and *Thunbergia*, and graminoids, *Trachypogon*, had lower FOO values of 8%. Despite the low FOO% values displayed by forb taxa identified by the *rbcL* marker, the *trn*L marker analysis showed the significance of forbs in the dietary composition, with Malvaceae emerging as the most frequently ingested plant family with a FOO of 92%. Ebenaceae and Poaceae, representing woody taxa and graminoids respectively, followed closely with 84% each (Fig. 2). Conversely, the forb family, Asteraceae was the least frequently occurring family in the dataset (FOO = 8%). Among the genera identified by the *trn*L marker, the woody taxa *Euclea* and *Senegalia* as well as the forb, *Ipomoea*, showed high frequencies, with FOO values of 84%, 72%, and 48%, respectively. Other genera belonging to the woody taxa were the genera *Ficus* and *Ehretia*, representing the least frequent dietary items with each recording a FOO of 12%.

In terms of forb abundance, the Welverdiend *rbc*L dataset showed lower abundance for genera like *Malvastrum* and *Thunbergia* compared to the dominance of woody taxa, but *trn*L identifications highlight the importance of forbs with Malvaceaea being the dominant family. This marked difference in datasets underscores the need for a dual-locus approach to fully capture the richness and abundance of forbs in cattle diets. The RRA calculations for the Welverdiend *rbc*L dataset (Fig. 3), showed that forbs/woody taxa were most abundant in 18 samples, followed by six samples where woody taxa were most abundant. Fabaceae, Combretaceae, and Ebenaceae were identified as the main food items across most samples. These families represent a variety of life forms, especially including woody taxa like *Albizia*, *Combretum*, *Euclea*, and *Diospyros*, indicating the significant contribution of shrubs and trees to the diets of Welverdiend cattle. However, variations from this overall pattern were observed for specific samples. Sample 16 exhibited a dominance of Malvaceae (forbs/non-woody shrubs), with

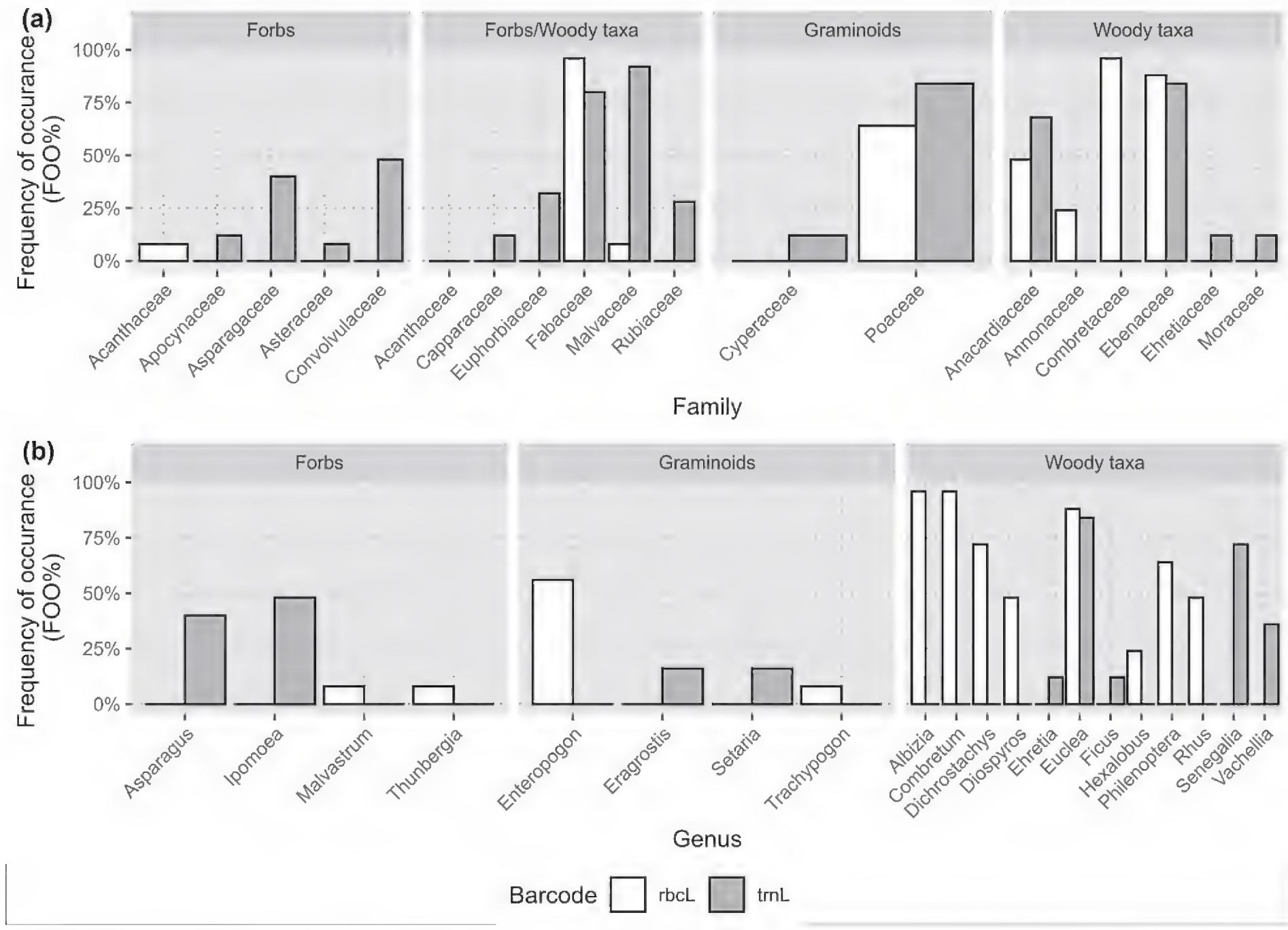

Figure 2. Bar chart showing the percentage of frequency of occurrence (FOO%) of the plant (**a**) families and (**b**) genera, identified as part of the diets of Welverdiend cattle, faceted according to their life form. White bars represent taxa identified with *rbc*L and grey bars represent those identified with *trn*L.

greater inclusion of Poaceae (graminoids) compared to other samples. Samples 22 and 23 stood out with Anacardiaceae (woody taxa) as the dominant food source. The genus *Rhus* (shrub with synonym *Searsia*) belonging to this family, was identified as an important food item in samples 17, 18, 19, 4, and specifically samples 22 and 23, where *Rhus* represented more than 75%.

Forbs/woody taxa were again identified as the dominant life form in most samples in the Welverdiend *trn*L dataset followed by woody taxa (Fig. 3). Malvaceae, composed of forbs and non-woody shrubs, was the most abundant family in many samples. Ebenaceae, composed of woody taxa such as *Euclea*, also played a significant role in the contribution of identified families in samples 6, 2, 3, and 21. While less abundant than the latter families, Convolvulaceae was also important across many samples and included the forb *Ipomoea*. Grasses (Poaceae), including *Eragrostis*, *Setaria*, and *Enteropogon*, served as a vital food item across samples, however, it did not dominate the dietary compositions; its relative abundance was lower compared to other families like Malvaceae and Ebenaceae in most samples (Fig. 3). Other forb taxa such as Asparagaceae, Rubiaceae, and Euphorbiaceae also contributed to cattle diets in smaller relative amounts. Furthermore, the analysis of genera identified for the Welverdiend *trn*L dataset revealed that *Euclea* (woody taxa) was the most abundant, followed by *Ipomoea* (forb), *Vachellia* (shrub/tree), *Senegalia* (shrub/tree), *Asparagus* (forb) and *Eragrostis* (grass) as well as *Ficus* (shrub/tree) in samples 18 and 23.

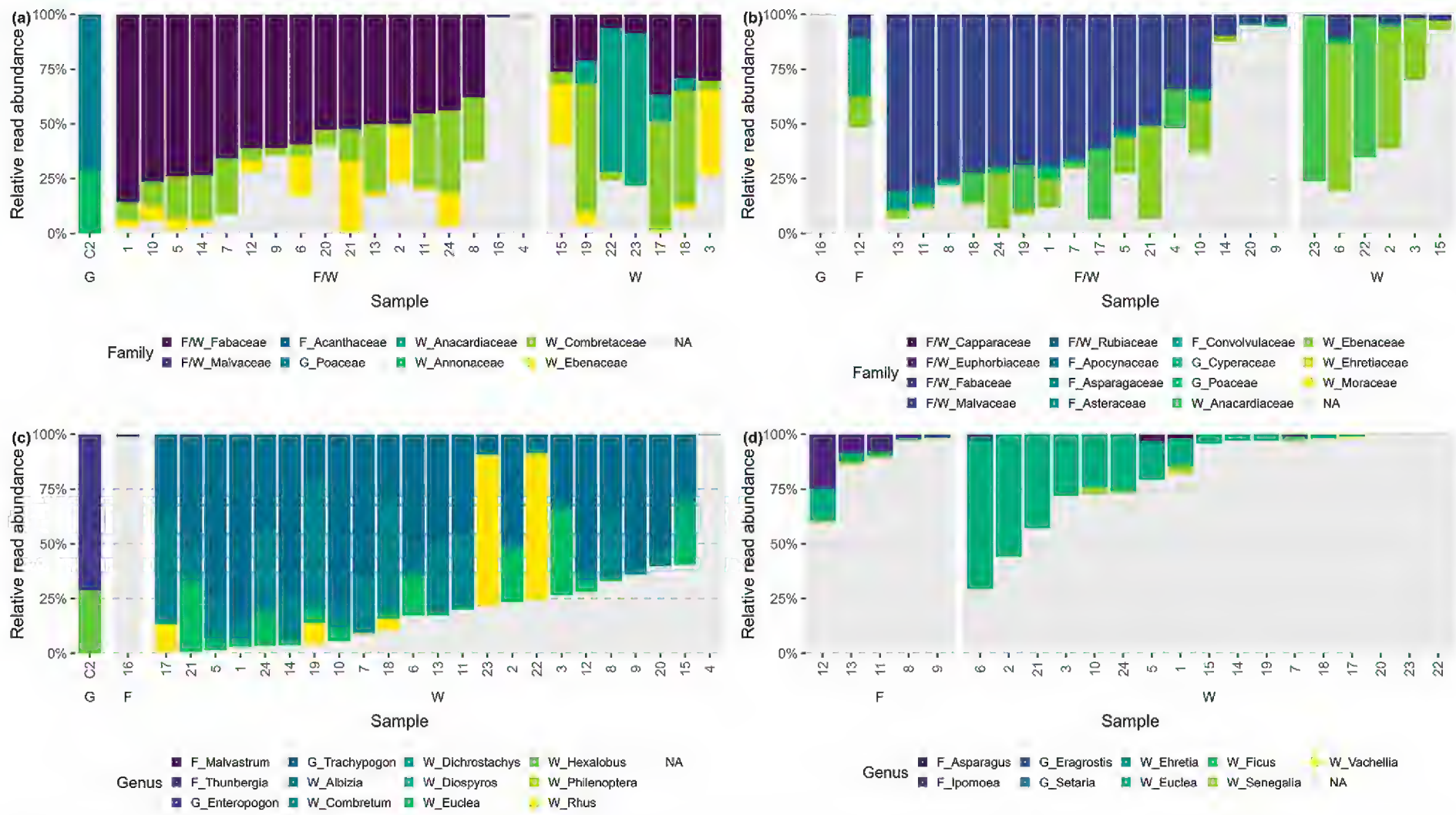

Figure 3. Relative read abundance (RRA) of families (**a, b**) and genera (**c, d**) identified with the *rbc*L (**a, c**) and *trn*L (**b, d**) barcoding marker from the Welverdiend dataset. The stacked bar plots the distribution of plant taxa in individual samples arranged according to their dominant life forms relative to the number of unassigned reads, with codes indicating graminoids (G), forbs (F), forbs/woody taxa (F/W) and woody taxa (W). For the *trn*L charts, the sample with no families identified (C2) and samples with no genera identified (4 and 16) were excluded from the visualisations.

At Syferkuil, a different pattern emerged, with graminoids and woody taxa dominating both marker datasets, but particularly so in *rbc*L (Fig. 4). Forb and woody taxa frequency was lower overall compared to Welverdiend, with graminoids dominating the cattle diets. This is evident from the *rbc*L dataset where Poaceae was the most frequently occurring family (88%). Fabaceae emerged as the second most frequently occurring family (75%), which comprises both forbs and woody taxa. This was followed by Rhamnaceae, containing woody taxa, with a 67% FOO. Conversely, the families representing forbs, Geraniaceae, and Solanaceae were the least frequently occurring families, with 8% each. *Tricholaena*, a graminoid, was the most frequently occurring genus (79%). This was followed by *Cenchrus* (graminoid) with 71%, and *Ziziphus* (woody taxa) with a 71% FOO. The least frequent genera were *Berchemia* (woody taxa), and the forbs *Monsonia* and *Solanum* with FOO values of 8%, each. In the *trn*L dataset (Fig. 4), the dominance of grasses was confirmed with Poaceae having 96% FOO. Like in the *rbc*L dataset, this was followed by Fabaceae (woody taxa/forbs) with 79% FOO. In the *trn*L dataset, Caryophyllaceae (forb) emerged as one of the most frequently utilised families with a FOO of 25%. Other families representing forbs such as Solanaceae, Lamiaceae, and Euphorbiaceae were the least frequent (8%). The genus *Setaria*, representative of graminoids, was the most common (88%), followed by *Vachellia* (woody taxa) with a 58% FOO. *Pollichia* and *Indigofera*, consisting of forbs each had a 25% FOO while other forbs such as *Leucas*, *Senna*, and *Solanum*, and a woody taxon, *Senegalia*, only had an 8% FOO.

Reinforcing the dominance of grasses in the diets of cattle at Syferkuil, the *rbc*L marker (Fig. 5) primarily identified graminoids across most samples, with

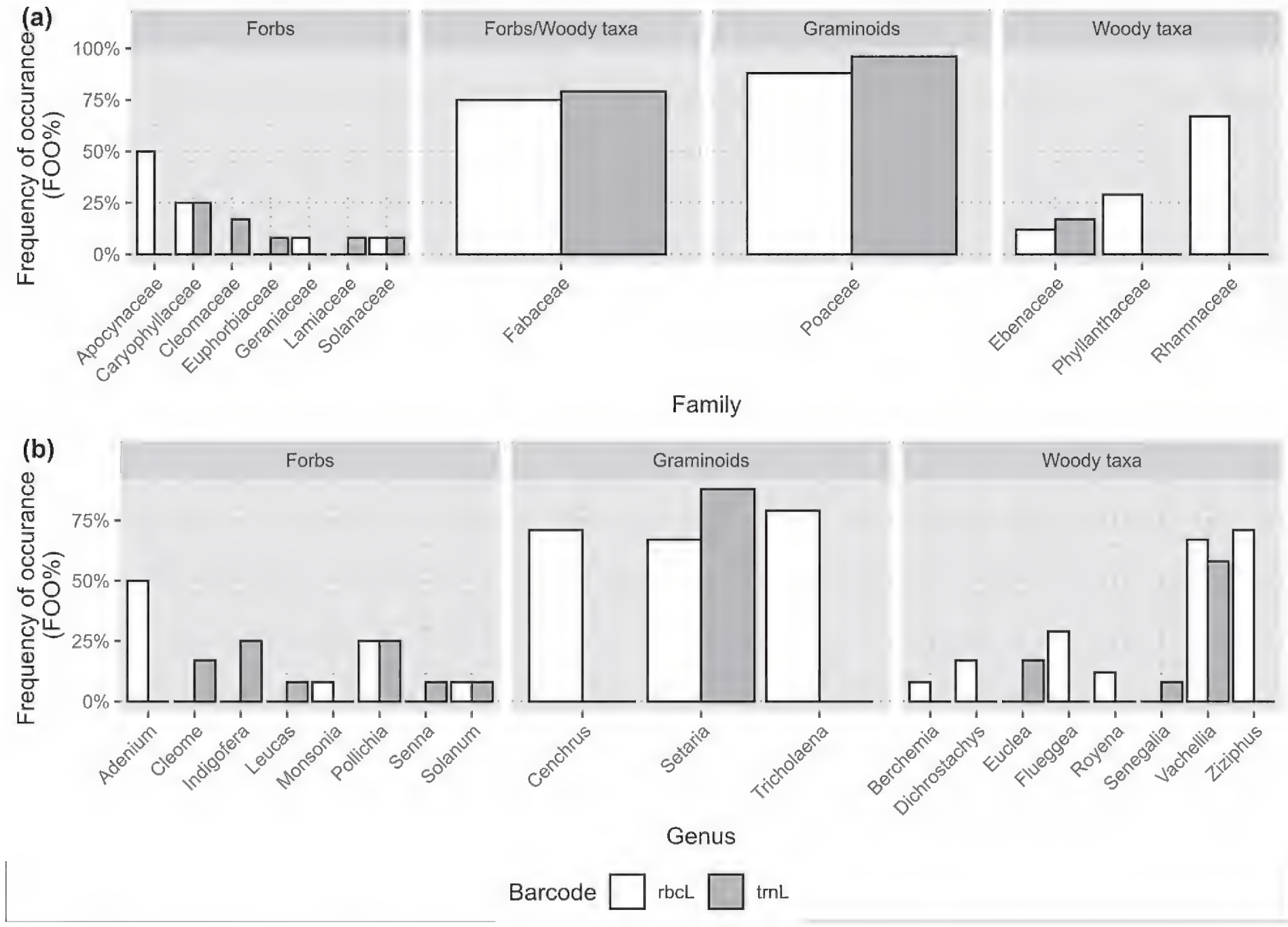

Figure 4. Bar chart showing the percentage of frequency of occurrence (FOO%) of the plant (**a**) families and (**b**) genera, identified as part of the diets of Syferkuil cattle, faceted according to their life form. White bars represent taxa identified with *rbc*L and grey bars represent those identified with *trn*L.

notable genera such as *Setaria*, *Cenchrus*, and *Tricholaena*. Woody taxa such as *Vachellia* (Fabaceae), *Flueggea* (Ebenaceae), and *Ziziphus* (Rhamnaceae) were also significant components. Forbs such as *Pollichia* showed some diversity across the samples, although in smaller abundances compared to other life forms. Forbs were identified as the dominant life form in 5 samples. The forb family, Caryophyllaceae, was especially abundant in sample 25 where it was the only family detected, and Solanaceae which appeared in small amounts in samples 33 and 46. Another important forb family was Apocynaceae which represented more than 90% of the identified plant families in sample 41. Following the patterns of grass-dominated diets at Sygerkuil, *trn*L RRA calculations showed that Poaceae was the dominant life form across 20 samples. Woody taxa (Fabaceae) also contributed a relatively large proportion (Fig. 5). Additionally, *Setaria* (grass) and *Vachellia* (shrub/tree) were identified as dominant genera in Syferkuil cattle diets with *Euclea* (shrub/tree), *Pollichia* (forb), *Cleome* (forb) and *Indigofera* (forb/shrub) in lower relative abundances. *Leucas* (forb) was also identified as an important food item in samples 28 and 47 and *Senna* (forb) in smaller proportions in samples 28 and 37.

In summary, the RRA analysis across both sites showed a predominant reliance on graminoids and woody taxa, with forbs contributing to the diets, but in smaller relative amounts. The cattle diets at Welverdiend showed a higher overall dominance of forb and woody taxa compared to those of Syferkuil,

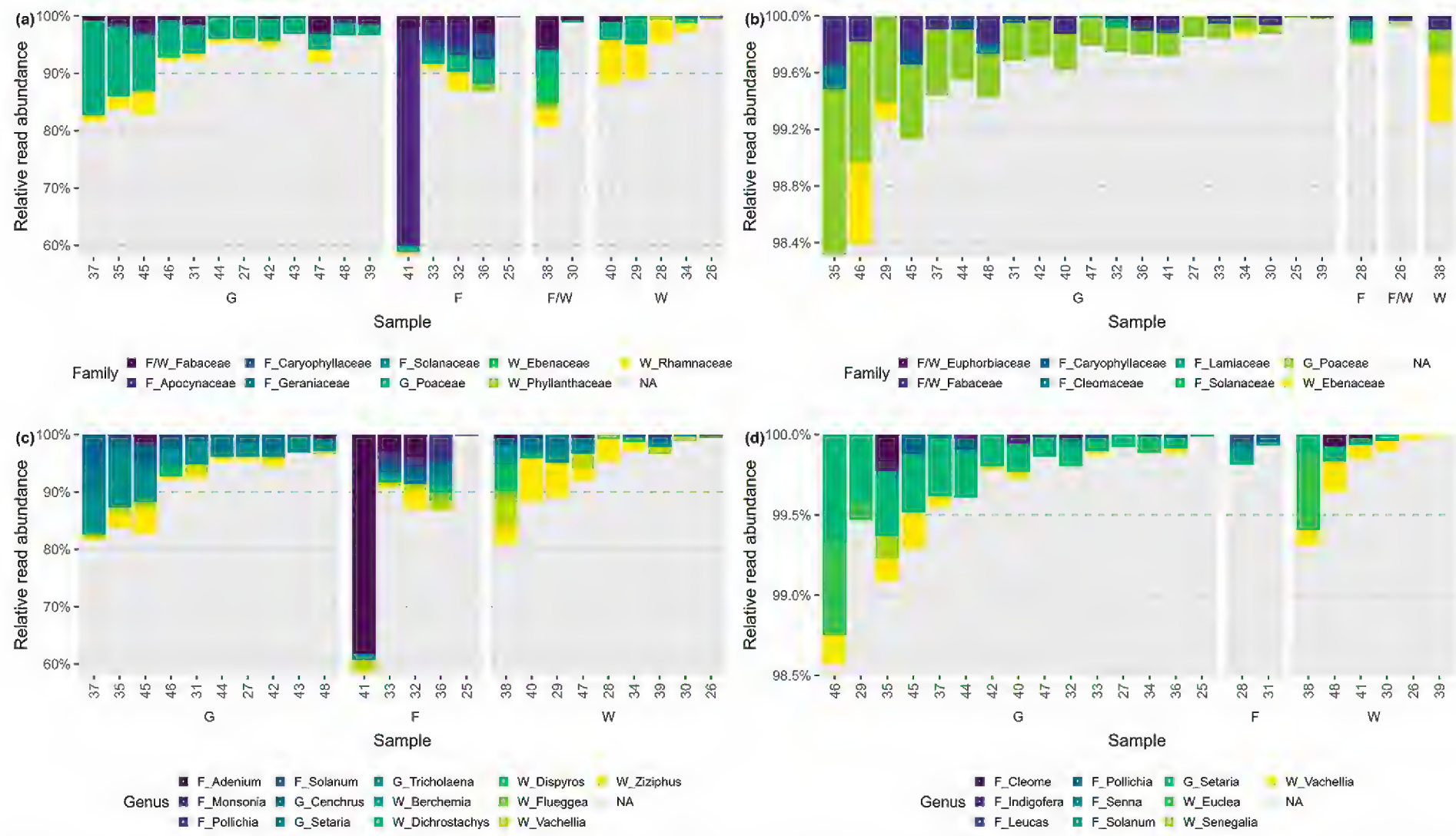

Figure 5. Relative read abundance (RRA) of families (**a, b**) and genera (**c, d**) identified with the *rbc*L (**a, c**) and *trn*L (**b, d**) barcoding marker from the Syferkuil dataset. The stacked bar plots the distribution of plant taxa in individual samples arranged according to their dominant life forms relative to the number of unassigned reads, with codes indicating graminoids (G), forbs (F), forbs/woody taxa (F/W) and woody taxa (W). For the *trn*L charts, sample C2 had no family or genus identifications and was excluded from the visualisations.

which relied more heavily on grasses. The use of both the *rbc*L and *trn*L markers further highlighted the differences in the identification of plant families and genera, with *trn*L providing more detailed insights into forb consumption. Despite these differences in resolution between loci, there was consistency across both datasets in the identification of key families such as Fabaceae, Ebenaceaea, and Poaceae. These families were prominent in both Welverdiend and Syferkuil. However, the broader diversity captured by *trn*L, especially forbs, emphasises the importance of using multiple markers to obtain a comprehensive view of the dietary composition of the cattle foraging in east semi-arid African savannas.

Discussion

A dual-locus barcoding approach

In this study, we determined the dietary composition of cattle foraging in two bioregions of the eastern semi-arid South African savanna and demonstrated that forbs and woody taxa make a substantial contribution to the overall diet of cattle during the dry periods. The taxonomies described using the different barcoding regions highlight the need for the use of a dual-locus identification system for plants (CBOL Plant Working Group 2009; Hollingsworth et al. 2011). The tendency of barcoding loci to fail or weakly amplify specific taxonomic groups (CBOL Plant Working Group 2009; Zhu et al. 2022) allows for the systematic underrepresentation of inefficiently amplified species and the omission

of food items. Using multiple loci to recreate diet profiles allows for mutual verification, which helps reveal such disparities (Camp et al. 2020). The use of the *rbc*L and *trn*L chloroplast genes resulted in complementing taxonomic identifications, facilitating the reconstruction of herbivore diets from an eastern semi-arid South African savanna (Figs 2, 4, Table 1). Frequently consumed food items in the diets of Welverdiend cattle, according to the *rbc*L and *trn*L marker, were from the Fabaceae, Combretaceae, Malvaceae, Ebenaceae and Anacardiaceae. The family Fabaceae, Poaceae and Ebenaceae also occurred in relatively greater proportions in the diets of these cattle according to the RRA graphs except for Malvaceae in the *rbc*L dataset, where it made up ~63% of the diet of one individual and Combretaceae, which the *trn*L barcode could not identify. A well-established reason for the differing success of plant barcoding loci is due to the completeness of barcode reference libraries, especially those containing representative barcode records from South Africa (Rattray et al. 2024). Dissimilarity in the recorded plant diversity identified by barcoding loci could be due to added information due to differing taxonomic breadths (Fahner et al. 2016), which may explain why Combretaceae was missing from identifications with the *trn*L dataset. This family was represented by 12 unique species in the combined reference libraries, of which only three were barcoded with *trn*L and only consisted of one reference sequence each (Botha et al. 2023). The *rbc*L barcode, however, was represented by all 12 species, of which nine were unique to this dataset, with two or more reference sequences each. Interestingly, although Malvaceae was one of the most frequently occurring families in the Welverdiend *trn*L dataset, it could not be identified to genus level, but it was in the *rbc*L dataset in the form of *Malvastrum* where it accounted for ~8% of the FOO and ~63% RRA in sample 16.

The high number of *trn*L ASVs left after filtering, combined with its greater phylogenetic distances and alpha diversity metrics, indicates that *trn*L is more robust than *rbc*L for capturing the diversity and phylogenetic structure of food items contained in herbivorous diets. Syferkuil exhibited the highest alpha diversity (Shannon index) and species richness for both markers, especially for *trn*L. This suggests that plant communities contained in cattle diets are more diverse compared to those at Welverdiend (Kruskal-Wallis test: $p < 0.05$). A substantial number of ASVs remained unassigned after comparison with reference libraries and could also not be assigned to plant species in the NCBI nt database. This indicates potential gaps in these reference sequence libraries or the presence of unknown or poorly characterised plant species in eastern semi-arid South African savannas, further emphasising the need for expanded reference databases for accurate metabarcoding analyses that will ideally facilitate multi-barcode approaches for plant identifications (Botha et al. 2023).

Except for the differences in families identified, the *rbc*L and *trn*L barcoding markers have different affinities for amplifying certain genera from a common family identified in the diets of cattle foraging in each study site. The taxa under consideration, as well as the marker used, have a significant impact on the success of species identification using DNA metabarcoding (De Vere et al. 2012; Mallott et al. 2018; Amandita et al. 2019) and dissimilarities in the identified taxa may increase at lower taxonomic levels to result in marker disagreement at genus level plant diversity (Yoccoz et al. 2012; Fahner et al. 2016). *Rbc*L and *trn*L has modest discriminatory power (Chase et al. 2005; Newmaster et

al. 2006; Taberlet et al. 2007; Hollingsworth et al. 2009; Valentini et al. 2009; Quéméré et al. 2013), but the observation of an overlapping identification between these markers can provide increased support for the findings (Kress and Erickson 2007; Fahner et al. 2016). In the Welverdiend study site, the family Fabaceae was represented by *Albizia*, *Dichrostachys* and *Philenoptera* as identified by the *rbc*L barcoding marker and *Senegalia* and *Vachellia* as identified by the *trn*L barcoding marker. For Syferkuil this same family was represented by *Dichrostachys* and *Vachellia* (shrubs and trees) by the *rbc*L marker and *Senegalia* (shrub/tree), *Vachellia* (shrub/tree), *Indigofera* (forb) and *Senna* (shrub/forb) by the *trn*L marker. The *rbc*L barcoding marker therefore exclusively identified the genus *Dichrostachys* across both datasets and the *trn*L marker exclusively identified *Senegalia*.

The identification of forbs

Identified food items across the studied sites were representatives from 25 different plant families, of which 16 included forbs as a potentially foraged life form (Table 1). Forb genera were frequently identified within the sampled datasets as foraged items at both Welverdiend and Syferkuil. Genera representing forb species that were identified for the Welverdiend cattle are *Malvastrum* and *Thunbergia* as identified by the *rbc*L marker and *Asparagus*, and *Ipomoea* identified with *trn*L. These genera (*Malvastrum* and *Thunbergia*) are not present in the *trn*L reference dataset and did not appear frequently across the sampled dataset with values of 8%. Forb genera were identified at a much higher frequency in the *trn*L dataset with values of 40% for *Asparagus* and 48% for *Ipomoea*. The *trn*L barcoding marker therefore could allow for the identification of the genera *Asparagus* and *Ipomoea* even though there were reference sequences available in the *rbc*L barcoding library. For the Syferkuil cattle the genera *Pollichia* (forb; FOO = 25%), *Adenium* (succulent shrub; FOO = 50%), *Solanum* (forb/shrub; FOO = 8.33%), and *Monsonia* (forb; FOO 8.33%) were included in their diets as identified by the *rbc*L marker, as well as *Solanum* (forb/shrub; FOO = 8.33%), *Cleome* (forb; FOO = 16.7%), *Indigofera* (forb; FOO = 25%), *Leucas* (forb; FOO = 8.33%) and *Pollichia* (forb; FOO = 25%) identified by the *trn*L marker. Only the genera of *Pollichia* and *Solanum* could be identified by both the barcoding markers, again indicating the complementary nature of this dual-locus identification system that aids in the determination of dietary profiles when reference sequences are not available for all barcodes (Taberlet et al. 2007) or when a barcode shows preferential amplification of certain plant taxa (Kartzinel et al. 2015; Mallott et al. 2018; Zhu et al. 2022).

Cattle diets during the dry season in a semi-arid African savanna

According to the results, forbs are foraged by a noticeable percentage of individuals in the sampled dataset, especially species from the genera *Asparagus* and *Ipomoea* at Welverdiend and *Adenium*, *Pollichia* and *Indigofera* at Syferkuil. Different forb community compositions and diversity across these two study sites (Mucina and Rutherford 2006) were confirmed by the RRA and FOO visualisations. Welverdiend had the highest frequency of occurrence of forb families (families including forbs/woody taxa) across the sampled datasets (i.e., higher

forb species richness in the diet), while Syferkuil results illustrated a greater relative abundance of the forbs identified in individual diets (i.e., lower richness, but higher relative abundance per species). This suggests that although the forbs occurring at Welverdiend were foraged upon by many members of the sampled dataset, they have not foraged heavily, and their diets still included a high abundance of other life forms. In contrast a higher preference for forbs was observed at Syferkuil. The overall dietary composition of cattle foraging in a semi-arid African savanna at the end of the dry season shows a transition from usual grass-based forage to a stronger reliance on forbs and leaves from woody taxa such as trees and shrubs.

The interaction between availability and selection determines the actual diet consumed by animals (Julian et al. 2024). When availability is high, animals can be more selective, potentially improving their nutritional intake. When availability is low, selection becomes constrained, and animals may be forced to consume less preferred or lower-quality foods. In Welverdiend, forbs and woody taxa from families such as Malvaceae and Ebenacea are the main parts of cattle diets. This indicates how crucial these life forms become when there is less grass available, or as shown by Craine (2021) and Craine et al. (2015), when graminoids have lower nutrient concentrations with warming climate and large mammalian herbivores rely on shrubs and forbs for their protein intake. This was also observed at Syferkuil, though grasses still dominate the cattle diets, consumption of woody plants and forbs from families such as Ebenaceae and Caryophyllaceae also occurs. This reliance on forbs and woody life forms in the dry season suggests that they have a flexible foraging strategy that allows them to survive through periods of drought while maintaining dietary diversity and a nutritional balance.

Recognising and understanding these dietary shifts in cattle diets can guide adaptive management practices and will ensure the sustainable use of forage in semi-arid landscapes while sustaining livestock productivity. To mitigate the impacts of nutrient stress imposed by increasing temperature and decreasing precipitation, managers should consider implementing rotational grazing, providing supplemental feeding when necessary, and conserving the diversity of not only grasses but of shrubs and forbs in grazing areas to support flexible foraging strategies during periods of scarcity in dryland landscapes (Samuels et al. 2016; Scasta et al. 2020; Julian et al. 2024).

While the analysis provides valuable insights into the foraging behaviour of these herbivores, it's important to consider potential biases inherent to the methodology. PCR bias may result in selective amplification of certain taxa, leading to an over- or underrepresentation of specific plant taxa. Furthermore, not all taxa that are present in the samples may have been successfully amplified and identified, potentially skewing the perceived diversity and richness of the cattle diets. Moreover, the faecal samples that were collected represent a "snapshot" of what an individual consumed immediately before sampling, rather than providing an overview of their long-term dietary habits. The authors recommend exploring methods to mitigate PCR bias and improve the representativeness of sampling by incorporating multiple sampling points over time. Additionally, integrating dietary observations or tracking methods alongside genetic analyses could offer a more holistic understanding of cattle diets and foraging behaviour.

Conclusion

The food items occurring in the diets of herbivores foraging in two bioregions of an eastern semi-arid South African savanna could be recovered and successfully taxonomically assigned using a dual primer approach and study-area specific reference libraries. The large number of unassigned reads retrieved by both markers is largely indicative of incomplete reference libraries and highlights the need for more complete and accurate multi-barcode records from South African savannas; not just to enhance the reconstruction of dietary profiles but to support other metabarcoding studies dealing with plant identifications from this region. This study shows the success of a dual-locus barcoding approach for the identification of plants in that the *rbc*L marker supplemented missing taxonomies for *trn*L and vice versa. The use of two plant barcoding regions was therefore necessary to retrieve a more accurate taxonomic resolution to reconstruct the diets of herbivores, especially important nutritional food items such as forbs that are seasonally selected for, but not abundant in savanna ecosystems.

The importance of forbs was showcased via RRA and FOO calculations of the diets of Welverdiend and Syferkuil cattle where they represented 64% of the identified plant families, although in lower abundance and occurring less frequently compared to trees, shrubs and graminoids. Key forb genera in Welverdiend included *Malvastrum*, *Asparagus*, and *Ipomoea* while *Adenium* and *Pollichia* were important for cattle at Syferkuil. Although grasses dominated the diets at Syferkuil, these cattle also consumed woody taxa and forbs, which were heavily foraged upon when encountered. In contrast, cattle in the communal rangeland at Welverdiend displayed a greater proportion of woody taxa and forbs in their diets compared to grasses. While these forbs and woody taxa were foraged upon by most members of the sampled community, they did not dominate individual diets as seen for Syferkuil cattle. This indicates a shift in dietary reliance at the end of the dry season, where forbs and woody taxa become crucial for Welverdiend cattle due to the scarcity of grasses, while forbs and woody taxa are actively foraged on at Syferkuil experimental farm for their nutritional benefits.

This study emphasises the flexibility in cattle foraging behaviour in the dry season when more forbs and woody-taxa are included in their grass-based diets, allowing them to adapt to maintain nutritional balances despite changing environmental conditions and resource availability. Such adaptability gives rise to the need for adaptive management strategies in semi-arid ecosystems to maintain livestock productivity. Furthermore, the preferences exhibited by cattle for different plant growth forms and taxa in semi-arid African savannas present an opportunity for future work to better understand the flexible forging behaviours of other large mammalian herbivores under the influence of drier climates. Further investigations of these flexible foraging strategies can provide valuable insights into the protection of wild herbivores, a better understanding of forb ecology, and the overall maintenance of ecosystem health.

Acknowledgements

We wish to thank the South African Environmental Observation Network (SAEON), Ndlovu Node and the University of Limpopo for providing logistical support during field data collection. We are grateful to NICD for the metabarcoding of samples.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

The North-West University Animal Production Sciences Research Ethics Committee approved the study as zero-risk and granted written permission for its completion. The ethics number is NWU-00177-18-A5.

Funding

We thank the North-West University for its support of this research. The research was also supported by the German-South African research project “South African Limpopo Landscapes Network” (SALLnet) funded by the German Federal Ministry of Education and Research (BMBF; Grant number: 01LL1802C).

Author contributions

Conceptualization: SB, AL, FS. Data curation: FS, SB, DB. Formal analysis: DB. Funding acquisition: AL, FS. Investigation: AL, FS, DB, KB, MVM, SB. Methodology: DB, MP, SB. Project administration: FS, KB, SB, AL. Resources: ZTHM, SB, FS, MA, AI, AL. Software: DB, MP. Supervision: SB, FS. Validation: SB, DB, FS, AL. Visualization: DB. Writing - original draft: DB, FS, SB. Writing - review and editing: AI, MVM, KB, AL, MA, MP, ZTHM.

Author ORCIDs

Danielle Botha https://orcid.org/0000-0002-7321-5382
Sandra Barnard https://orcid.org/0000-0002-9299-2769
Morné du Plessis https://orcid.org/0000-0003-4154-7830
Mushal Allam https://orcid.org/0000-0002-9875-6716
Kai Behn https://orcid.org/0000-0003-2598-9575
Arshad Ismail https://orcid.org/0000-0003-4672-5915
Anja Linstädter https://orcid.org/0000-0003-0038-9557
Malesela Vincent Mokoka https://orcid.org/0000-0002-1026-2982
Zamantungwa Thobeka Happiness Mnisi https://orcid.org/0000-0002-7923-5077
Frances Siebert https://orcid.org/0000-0001-5549-8211

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

References

Acocks JPH (1988) Veld types of South Africa (3rd edn). Memoirs of the Botanical Survey of South Africa 57: 1–146.

Amandita FY, Rembold K, Vornam B, Rahayu S, Siregar IZ, Kreft H, Finkeldey R (2019) DNA barcoding of flowering plants in Sumatra, Indonesia. Ecology and Evolution 9(4): 1858–1868. https://doi.org/10.1002/ece3.4875

Aylward CM, Statham MJ, Barthman-Thompson L, Kelt DA, Sacks BN (2022) Dietary characterization of the endangered salt marsh harvest mouse and sympatric ro-

dents using DNA metabarcoding. Ecology and Evolution 12(7): e9121. https://doi.org/10.1002/ece3.9121

Ayugi B, Eresanya EO, Onyango AO, Ogou FK, Okoro EC, Okoye CO, Anoruo CM, Dike VN, Ashiru OR, Daramola MT, Mumo R, Ongoma V (2022) Review of meteorological drought in Africa: Historical trends, impacts, mitigation measures, and prospects. Pure and Applied Geophysics 179(4): 1365–1386. https://doi.org/10.1007/s00024-022-02988-z

Bhattacharyya S, Dawson DA, Hipperson H, Ishtiaq F (2018) Diet rich in C3 plants reveals the sensitivity of an alpine mammal to climate change. Molecular Ecology 28(2): 250–265. https://doi.org/10.1111/mec.14842

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature Methods 10(1): 57–59. https://doi.org/10.1038/nmeth.2276

Botha D (2023) A comparison of bioinformatic workflows for herbivore dietary analysis using a DNA metabarcoding approach. [Masters Dissertation, North-West University]. Natural and Agricultural Sciences Collection.

Botha D, Du Plessis M, Siebert F, Barnard S (2023) Introducing an rbcL and a trnL reference library to aid in the metabarcoding analysis of foraged plants from two semi-arid eastern South African savanna bioregions. PLoS ONE 18(5): e0286144–e0286144. https://doi.org/10.1371/journal.pone.0286144

Buglione M, Maselli V, Rippa D, de Filippo G, Trapanese M, Fulgione D (2018) A pilot study on the application of DNA metabarcoding for non-invasive diet analysis in the Italian hare. Mammalian Biology 88: 31–42. https://doi.org/10.1016/j.mambio.2017.10.010

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13(7): 581–583. https://doi.org/10.1038/nmeth.3869

Camp A, Croxford AE, Ford CS, Baumann U, Clements PR, Hiendleder S, Woolford L, Netzel G, Boardman WSJ, Fletcher MT, Wilkinson MJ (2020) Dual-locus DNA metabarcoding reveals southern hairy-nosed wombats (*Lasiorhinus latifrons* Owen) have a summer diet dominated by toxic invasive plants. PLoS ONE 15(3): e0229390. https://doi.org/10.1371/journal.pone.0229390

CBOL Plant Working Group (2009) A DNA barcode for land plants. Proceedings of the National Academy of Sciences 106(31): 12794–12797. https://doi.org/10.1073/pnas.0905845106

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V (2005) Land plants and DNA barcodes: Short-term and long-term goals. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 360(1462): 1889–1895. https://doi.org/10.1098/rstb.2005.1720

Chivangulula FM, Amraoui M, Pereira MG (2023) The drought regime in Southern Africa: A systematic review. Climate (Basel) 11(7): 147. https://doi.org/10.3390/cli11070147

Coissac E, Hollingsworth PM, Lavergne S, Taberlet P (2016) From barcodes to genomes: Extending the concept of DNA barcoding. Molecular Ecology 25(7): 1423–1428. https://doi.org/10.1111/mec.13549

Craine JM (2021) Seasonal patterns of bison diet across climate gradients in North America. Scientific Reports 11(1): 6829. https://doi.org/10.1038/s41598-021-86260-9

Craine JM, Towne G, Miller M, Fierer N (2015) Climatic warming and the future of bison as grazers. Scientific Reports 5(1): 16738. https://doi.org/10.1038/srep16738

De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, Taberlet P (2014) DNA metabarcoding multiplexing and validation of data accuracy for diet assessment:

Application to omnivorous diet. Molecular Ecology Resources 14(2): 306–323. https://doi.org/10.1111/1755-0998.12188

De Groot GA, During HJ, Maas JW, Schneider H, Vogel JC, Erkens RHJ (2011) Use of rbcL and trnL-F as a two-locus DNA barcode for identification of NW-European ferns: An ecological perspective. PLoS ONE 6(1): e16371. https://doi.org/10.1371/journal.pone.0016371

De Vere N, Rich TCG, Ford CR, Trinder SA, Long C, Moore CW, Satterthwaite D, Davies H, Allainguillaume J, Ronca S, Tatarinova T, Garbett H, Walker K, Wilkinson MJ (2012) DNA barcoding the native flowering plants and conifers of Wales. PLoS ONE 7(6): e37945. https://doi.org/10.1371/journal.pone.0037945

Deagle BE, Eveson JP, Jarman SN (2006) Quantification of damage in DNA recovered from highly degraded samples – a case study on DNA in faeces. Frontiers in Zoology 3(1): 11. https://doi.org/10.1186/1742-9994-3-11

Deagle BE, Chiaradia A, McInnes J, Jarman SN (2010) Pyrosequencing faecal DNA to determine diet of little penguins: Is what goes in what comes out? Conservation Genetics 11(5): 2039–2048. https://doi.org/10.1007/s10592-010-0096-6

Department of Agriculture Land Reform and Rural Development (2022) Statistics and economic publications and reports: National livestock statistics May 2022. https://old.dalrrd.gov.za/Home/Crop-Estimates/Statistical-Information/Livestock

Distel RA, Arroquy JI, Lagrange S, Villalba JJ (2020) Designing diverse agricultural pastures for improving ruminant production systems. Frontiers in Sustainable Food Systems 4: 596869. https://doi.org/10.3389/fsufs.2020.596869

Fahner NA, Shokralla S, Baird DJ, Hajibabaei M (2016) Large-scale monitoring of plants through environmental DNA metabarcoding of soil: Recovery, resolution, and annotation of four dna markers. PLoS ONE 11(6): e0157505. https://doi.org/10.1371/journal.pone.0157505

Ferri G, Corradini B, Ferrari F, Santunione AL, Palazzoli F, Alu' M (2015) Forensic botany II, DNA barcode for land plants: Which markers after the international agreement? Forensic Science International. Genetics 15: 131–136. https://doi.org/10.1016/j.fsigen.2014.10.005

Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, Hansen AJ (2017) Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature Communications 8(1): 1188. https://doi.org/10.1038/s41467-017-01312-x

Garcia E, Siles P, Eash L, Van Der Hoek R, Kearney SP, Smukler SM, Fonte SJ (2018) Participatory evaluation of improved grasses and forage legumes for smallholder livestock production in central America. Experimental Agriculture 55(5): 776–792. https://doi.org/10.1017/S0014479718000364

García-Robledo C, Erickson DL, Staines CL, Erwin TL, Kress WJ (2013) Tropical plant–herbivore networks: Reconstructing species interactions using DNA barcodes. PLoS ONE 8(1): e52967. https://doi.org/10.1371/journal.pone.0052967

Gebremedhin B, Flagstad Ø, Bekele A, Chala D, Bakkestuen V, Boessenkool S, Popp M, Gussarova G, Schrøder-Nielsen A, Nemomissa S, Brochmann C, Stenseth NC, Epp LS (2016) DNA metabarcoding reveals diet overlap between the endangered Walia Ibex and domestic goats - implications for conservation. PLoS ONE 11(7): e0159133. https://doi.org/10.1371/journal.pone.0159133

Gerrish J (2004) Management-intensive grazing: the grassroots of grass farming. Green Park Press, Purvis, Miss.

Gill BA, Musili PM, Kurukura S, Hassan AA, Goheen JR, Kress WJ, Kuzmina M, Pringle RM, Kartzinel TR (2019) Plant DNA-barcode library and community phylogeny for

a semi-arid East African savanna. Molecular Ecology Resources 19(4): 838–846. https://doi.org/10.1111/1755-0998.13001

Goldberg AR, Conway CJ, Tank DC, Andrews KR, Gour DS, Waits LP (2020) Diet of a rare herbivore based on DNA metabarcoding of feces: Selection, seasonality, and survival. Ecology and Evolution 10(14): 7627–7643. https://doi.org/10.1002/ece3.6488

Hajibabaei M, Baird DJ, Fahner NA, Beiko R, Golding GB (2016) A new way to contemplate Darwin's tangled bank: How DNA barcodes are reconnecting biodiversity science and biomonitoring. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 371(1702): 20150330. https://doi.org/10.1098/rstb.2015.0330

Hempson GP, Archibald S, Bond WJ (2017) The consequences of replacing wildlife with livestock in Africa. Scientific Reports 7(1): 17196. https://doi.org/10.1038/s41598-017-17348-4

Higgins SI, Shackleton CM, Robinson ER (1999) Changes in woody community structure and composition under constrasting landuse systems in a semi-arid savanna, South Africa. Journal of Biogeography 26(3): 619–627. https://doi.org/10.1046/j.1365-2699.1999.t01-1-00317.x

Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, Chase MW, Cowan RS, Erickson DL, Fazekas AJ, Graham SW, James KE, Kim K-J, Kress WJ, Schneider H, van AlphenStahl J, Barrett SCH, van den Berg C, Bogarin D, Burgess KS, Cameron KM, Carine M, Chacón J, Clark A, Clarkson JJ, Conrad F, Devey DS, Ford CS, Hedderson TAJ, Hollingsworth ML, Husband BC, Kelly LJ, Kesanakurti PR, Kim JS, Kim Y-D, Lahaye R, Lee H-L, Long DG, Madriñán S, Maurin O, Meusnier I, Newmaster SG, Park C-W, Percy DM, Petersen G, Richardson JE, Salazar GA, Savolainen V, Seberg O, Wilkinson MJ, Yi D-K, Little DP, CBOL Plant Working Group (2009) A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America 106(31): 12794–12797. https://doi.org/10.1073/pnas.0905845106

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PLoS ONE 6(5): e19254. https://doi.org/10.1371/journal.pone.0019254

Iwanowicz DD, Vandergast AG, Cornman RS, Adams CR, Kohn JR, Fisher RN, Brehme CS (2016) Metabarcoding of fecal samples to determine herbivore diets: A case study of the endangered Pacific Pocket Mouse. PLoS ONE 11(11): e0165366. https://doi.org/10.1371/journal.pone.0165366

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL (2008) NCBI BLAST: A better web interface. Nucleic Acids Research 36(Web Server): W5–W9. https://doi.org/10.1093/nar/gkn201

Jorns TR, Scasta JD, Derner JD, Augustine DJ, Porensky LM, Raynor EJ, CARM Stakeholder Group (2023) Adaptive multi-paddock grazing management reduces diet quality of yearling cattle in shortgrass steppe. The Rangeland Journal 45(4): 160–172. https://doi.org/10.1071/RJ23047

Julian AAM, Scasta JD, Stewart WC (2024) Sheep winter diets parameterized with fecal DNA metabarcoding and forage sampling informs mineral nutrition management. Rangeland Ecology and Management 94: 168–177. https://doi.org/10.1016/j.rama.2024.03.003

Kartzinel TR, Chen PA, Coverdale TC, Erickson DL, Kress WJ, Kuzmina ML, Rubenstein DI, Wang W, Pringle RM (2015) DNA metabarcoding illuminates dietary niche partitioning by African large herbivores. Proceedings of the National Academy of Sciences of the United States of America 112(26): 8019–8024. https://doi.org/10.1073/pnas.1503283112

Katikati A, Fourie PJ (2019) Improving management practices of emerging cattle farmers in selected areas of the Eastern Cape Province: The role of agricultural extension. South African Journal of Agricultural Extension 47(1). https://doi.org/10.17159/2413-3221/2019/v47n1a492 [SAJAE]

Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution 30(4): 772–780. https://doi.org/10.1093/molbev/mst010

Kleppel GS (2020) Do Differences in Livestock Management Practices Influence Environmental Impacts? Frontiers in Sustainable Food Systems 4: 141. https://doi.org/10.3389/fsufs.2020.00141

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: The coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS ONE 2(6): e508. https://doi.org/10.1371/journal.pone.0000508

Kress WJ, García-Robledo C, Uriarte M, Erickson DL (2015) DNA barcodes for ecology, evolution, and conservation. Trends in Ecology & Evolution 30(1): 25–35. https://doi.org/10.1016/j.tree.2014.10.008

Lee T, Alemseged Y, Mitchell A (2018) Dropping Hints: Estimating the diets of livestock in rangelands using DNA metabarcoding of faeces. Metabarcoding and Metagenomics 2: e22467. https://doi.org/10.3897/mbmg.2.22467

Lemaire G, Hodgson J, Chabbi A (Eds) (2011) Grassland productivity and ecosystem services. CABI, Wallingford, Oxfordshire; Cambridge, Ma. https://doi.org/10.1079/9781845938093.0000

Levin RA, Wagner WL, Hoch PC, Nepokroeff M, Pires JC, Zimmer EA, Sytsma KJ (2003) Family-level relationships of Onagraceae based on chloroplast rbc L and ndh F data. American Journal of Botany 90(1): 107–115. https://doi.org/10.3732/ajb.90.1.107

Linstädter A, Kemmerling B, Baumann G, Kirscht H (2013) The importance of being reliable – Local ecological knowledge and management of forage plants in a dryland pastoral system (Morocco). Journal of Arid Environments 95: 30–40. https://doi.org/10.1016/j.jaridenv.2013.03.008

Linstädter A, Schellberg J, Brüser K, Moreno García CA, Oomen RJ, du Preez CC, Ruppert JC, Ewert F (2014) Are there consistent grazing indicators in drylands? Testing plant functional types of various complexity in South Africa's grassland and savanna biomes. PLoS ONE 9(8): e104672. https://doi.org/10.1371/journal.pone.0104672

Low AB, Rebelo AG (1998) Vegetation of South Africa, Lesotho and Swaziland. Department of Environmental Affairs and Tourism, Pretoria.

Mallott EK, Garber PA, Malhi RS (2018) trnL outperforms rbcL as a DNA metabarcoding marker when compared with the observed plant component of the diet of wild white-faced capuchins (*Cebus capucinus*, Primates). PLoS ONE 13(6): e0199556. https://doi.org/10.1371/journal.pone.0199556

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal 17(1): 10. https://doi.org/10.14806/ej.17.1.200

Matiz-Ceron L, Reyes A, Anzola J (2022) Taxonomical evaluation of plant chloroplastic markers by Bayesian classifier. Frontiers in Plant Science 12: 782663. https://doi.org/10.3389/fpls.2021.782663

Matsika R, Erasmus BFN, Twine WC (2013) A tale of two villages: Assessing the dynamics of fuelwood supply in communal landscapes in South Africa. Environmental Conservation 40(1): 71–83. https://doi.org/10.1017/S0376892912000264

McMurdie PJ, Holmes S (2013) phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8(4): e61217. https://doi.org/10.1371/journal.pone.0061217

Monkwe TR, Gxasheka M, Gunya B (2023) Challenges and perception of communal farmers on cattle production in Ga- Matlala, Limpopo Province, South Africa. Heliyon 9(3): e14190. https://doi.org/10.1016/j.heliyon.2023.e14190

Mucina L, Rutherford MC South (2006) The vegetation of South Africa, Lesotho and Swaziland. Sterlitzia 19. South African National Biodiversity Institute, Pretoria.

Mudzengi CP, Dahwa E, Kapembeza CS (2020) Livestock Feeds and Feeding in Semi-Arid Areas of Southern Africa. IntechOpen eBooks, 156 pp. https://doi.org/10.5772/intechopen.90109

Munjonji L, Ayisi KK, Mudongo EI, Mafeo TP, Behn K, Mokoka MV, Linstädter A (2020) Disentangling Drought and Grazing Effects on Soil Carbon Stocks and CO2 Fluxes in a Semi-Arid African Savanna. Frontiers in Environmental Science 8: 590665. https://doi.org/10.3389/fenvs.2020.590665

Mysterud A, Austrheim G (2016) The role of individual traits and environmental factors for diet composition of sheep. PLoS ONE 11(1): e0146217. https://doi.org/10.1371/journal.pone.0146217

Newmaster SG, Fazekas AJ, Ragupathy S (2006) DNA barcoding in land plants: Evaluation of rbcL in a multigene tiered approach. Canadian Journal of Botany 84(3): 335–341. https://doi.org/10.1139/b06-047

Nichols RV, Åkesson M, Kjellander P (2016) Diet assessment based on rumen contents: A comparison between DNA metabarcoding and macroscopy. PLoS ONE 11(6): e0157977. https://doi.org/10.1371/journal.pone.0157977

Odadi WO, Karachi MK, Abdulrazak SA, Young TP (2013) Protein supplementation reduces non-grass foraging by primary grazer. Ecological Applications 23(2): 455–463. https://doi.org/10.1890/12-0878.1

Odadi WO, Young TP, Okeyo-Owuor JB (2007) Effects of wildlife on cattle diets in Laikipia Rangeland, Kenya. Rangeland Ecology and Management 60(2): 179–185. https://doi.org/10.2111/05-044R3.1

Pansu J, Guyton JA, Potter AB, Atkins JL, Daskin JH, Wursten B, Kartzinel TR, Pringle RM (2018) Trophic ecology of large herbivores in a reassembling African ecosystem. Journal of Ecology 107(3): 1355–1376. https://doi.org/10.1111/1365-2745.13113

Pansu J, Hutchinson MC, Anderson TM, Te Beest M, Begg C, Begg KS, Bonin A, Chamaillé-Jammes S, Coissac É, Cromsigt JPGM, Demmel MY, Donaldson JE, Guyton JA, Hansen C, Imakando CI, Iqbal A, Kalima DF, Kerley GIH, Kurukura S (2022) The generality of cryptic dietary niche differences in diverse large-herbivore assemblages. Proceedings of the National Academy of Sciences of the United States of America 119(35): e2204400119. https://doi.org/10.1073/pnas.2204400119

Paradis E, Schliep K (2019) ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics (Oxford, England) 35(3): 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pauvert C, Buée M, Laval V, Edel-Hermann V, Fauchery L, Gautier A, Lesur I, Vallance J, Vacher C (2019) Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. Fungal Ecology 41: 23–33. https://doi.org/10.1016/j.funeco.2019.03.005

Phefadu KC, Kutu FR (2016) Evaluation of spatial variability of soil physico-chemical characteristics on rhodic ferralsol at the Syferkuil Experimental Farm of University of Limpopo, South Africa. Journal of Agricultural Science 8: 92. https://doi.org/10.5539/jas.v8n10p92

Quéméré E, Hibert F, Miquel C, Lhuillier E, Rasolondraibe E, Champeau J, Rabarivola C, Nusbaumer L, Chatelain C, Gautier L, Ranirison P, Crouau-Roy B, Taberlet P, Chikhi L (2013) A DNA Metabarcoding Study of a Primate Dietary Diversity and Plasticity across Its Entire Fragmented Range. PLoS ONE 8(3): e58971. https://doi.org/10.1371/journal.pone.0058971

R Core Team (2023) R: A language and environment for statistical computing. R-project.org. http://www.r-project.org/

Rattray RD, Stewart RD, Niemann HJ, Olaniyan OD, Van der Bank M (2024) Leafing through genetic barcodes: An assessment of 14 years of plant DNA barcoding in South Africa. South African Journal of Botany 172: 474–487. https://doi.org/10.1016/j.sajb.2024.07.055

Rivero MJ, Grau-Campanario P, Mullan S, Held SDE, Stokes JE, Lee MRF, Cardenas LM (2021) Factors affecting site use preference if grazing cattle studied from 2000 to 2020 through GPS tracking: A review. Sensors (Basel) 21(8): 2696. https://doi.org/10.3390/s21082696

Samuels I, Cupido C, Swarts MB, Palmer AR, Paulse JW (2016) Feeding ecology of four livestock species under different management in a semi-arid pastoral system in South Africa. African Journal of Range & Forage Science 33(1): 1–9. https://doi.org/10.2989/10220119.2015.1029972

Scasta JD, Jorns T, Derner JD, Lake S, Augustine DJ, Windh JL, Smith TL (2019) Validation of DNA metabarcoding of fecal samples using cattle fed known rations. Animal Feed Science and Technology 255: 114219. https://doi.org/10.1016/j.anifeedsci.2019.114219

Scasta JD, Jorns T, Derner JD, Stam B, McClaren M, Galkins G, Stewart W (2020) Technical note: Toxic plants in sheep diets grazing extensive landscapes: Insights from fecal DNA metabarcoding. Livestock Science 236: 104002. https://doi.org/10.1016/j.livsci.2020.104002

Scheiter S, Pfeiffer M, Behn K, Ayisi K, Siebert F, Linstädter A (2024) Managing Southern African Rangeland Systems in the Face of Drought: a Synthesis of Observation, Experimentation and Modeling for Policy and Decision Support. In: Von Maltitz G, Midgley GF, Veitch J, Brümmer C, Rötter RP, Viehberg F, Veste M (Eds) Sustainability of Southern African Ecosystems under Global Change. Ecological Studies, Vol 248. Springer, 439–470. https://doi.org/10.1007/978-3-031-10948-5_16

Ter Schure ATM, Pillai AAS, Thorbek L, Bhavani Shankar M, Puri R, Ravikanth G, Boer HJ, Boessenkool S (2020) eDNA metabarcoding reveals dietary niche overlap among herbivores in an Indian wildlife sanctuary. Environmental DNA 3(3): 681–696. https://doi.org/10.1002/edn3.168

Shackleton CM (2000) Comparison of plant diversity in protected and communal lands in the Bushbuckridge lowveld savanna, South Africa. Biological Conservation 94(3): 273–285. https://doi.org/10.1016/S0006-3207(00)00001-X

Shackleton SE, Shackleton CM, Netshiluvhi TR, Geach BS, Ballance A, Fairbanks DHK (2002) Use patterns and value of savanna resources in three rural villages in South Africa. Economic Botany 56(2): 130–146. https://doi.org/10.1663/0013-0001(2002)056[0130:UPAVOS]2.0.CO;2

Siebert F, Scogings P (2015) Browsing intensity of herbaceous forbs across a semi-arid savanna catenal sequence. South African Journal of Botany 100: 69–74. https://doi.org/10.1016/j.sajb.2015.05.007

Siebert F, Klem J, Van Coller H (2020) Forb community responses to an extensive drought in two contrasting land-use types of a semi-arid Lowveld savanna. African

Journal of Range & Forage Science 37(1): 53–64. https://doi.org/10.2989/10220119.2020.1726464

Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, Brysting AK, Sønstebø JH, Ims RA, Yoccoz NG, Taberlet P (2009) Analysing diet of small herbivores: The efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. Frontiers in Zoology 6(1): 16. https://doi.org/10.1186/1742-9994-6-16

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17(5): 1105–1109. https://doi.org/10.1007/BF00037152

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research 35(3): e14–e14. https://doi.org/10.1093/nar/gkl938

Tan G, Opitz L, Schlapbach R, Rehrauer H (2019) Long fragments achieve lower base quality in Illumina paired-end sequencing. Scientific Reports 9(1): 2856. https://doi.org/10.1038/s41598-019-39076-7

Thongjued K, Garcia K, Scott D, Gonthier DJ, Dupuis JR (2024) DNA metabarcoding diet analysis in a generalist omnivore: Feeding trials reveal the efficacy of extraction kits and a multi-locus approach for identifying diverse diets. Integrative Zoology 12(5): 790–806. https://doi.org/10.1111/1749-4877.12806

Trollope W, van Wilgen B, Trollope LA, Govender N, Potgieter AL (2014) The long-term effect of fire and grazing by wildlife on range condition in moist and arid savannas in the Kruger National Park. African Journal of Range & Forage Science 31(3): 199–208. https://doi.org/10.2989/10220119.2014.884511

Valentini A, Miquel C, Nawaz MA, Bellemain E, Coissac E, Pompanon F, Gielly L, Cruaud C, Nascetti G, Wincker P, Swenson J, Taberlet P (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: The trnL approach. Molecular Ecology Resources 9(1): 51–60. https://doi.org/10.1111/j.1755-0998.2008.02352.x

Villalba JJ, Provenza FD, Catanese F, Distel RA (2015) Understanding and manipulating diet choice in grazing animals. Animal Production Science 55(3): 261–271. https://doi.org/10.1071/AN14449

Wickham H (2016) Ggplot2: Elegant Graphics For Data Analysis. Springer-Verlag, New York. https://doi.org/10.1007/978-3-319-24277-4_9

Yang C, Wang X, Miller JA, de Blécourt M, Ji Y, Yang C, Harrison RD, Yu DW (2014) Using metabarcoding to ask if easily collected soil and leaf-litter samples can be used as a general biodiversity indicator. Ecological Indicators 46: 379–389. https://doi.org/10.1016/j.ecolind.2014.06.028

Yoccoz NG, Bråthen KA, Gielly L, Haile J, Edwards ME, Goslar T, Von Stedingk H, Brysting AK, Coissac E, Pompanon F, Sønstebø JH, Miqual C, Valentini A, De Bello F, Chave J, Thuiller W, Wincker P, Cruad C, Gavory F, Rasmussen M (2012) DNA from soil mirrors plant taxonomic and growth form diversity. Molecular Ecology 21(15): 3647–3655. https://doi.org/10.1111/j.1365-294X.2012.05545.x

Yu J, Xue J-H, Zhou S-L (2011) New universal matK primers for DNA barcoding angiosperms. Journal of Systematics and Evolution 49(3): 176–181. https://doi.org/10.1111/j.1759-6831.2011.00134.x

Zhu S, Liu Q, Qiu S, Dai J, Gao X (2022) DNA barcoding: An efficient technology to authenticate plant species of traditional Chinese medicine and recent advances. Chinese Medicine 17(1): 112. https://doi.org/10.1186/s13020-022-00655-y

Supplementary material 1

The sequences of the forward and reverse primers of the barcodes with their Illumina overhang adapters amplified from herbivorous faeces in this study

Authors: Danielle Botha, Sandra Barnard, Morné du Plessis, Mushal Allam, Kai Behn, Arshad Ismail, Anja Linstädter, Malesela Vincent Mokoka, Zamantungwa Thobeka Happiness Mnisi, Frances Siebert

Data type: xlsx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.127959.suppl1

Supplementary material 2

Overview of the main steps followed in the materials and methods section for the mock community analysis conducted in Botha (2023)

Authors: Danielle Botha, Sandra Barnard, Morné du Plessis, Mushal Allam, Kai Behn, Arshad Ismail, Anja Linstädter, Malesela Vincent Mokoka, Zamantungwa Thobeka Happiness Mnisi, Frances Siebert

Data type: png

Explanation note: * An external quality filter was added to the VSEARCH approach since this workflow does not incorporate one in its various steps. ** “OTU” in this diagram may refer to operational taxonomic units as retrieved from the clustering of sequences based on 97% similarity in the VSEARCH approach, amplicon sequence variants retrieved by the inference of an error model in the DADA2 approach or grouping of strictly identical sequences as in the OBITOOLS approach.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.127959.suppl2

Supplementary material 3

Bar plots generated from the taxonomic assignments of the Welverdiend rbcL (*rbcL*_WV) and trnL dataset (*trnL*_WV)

Authors: Danielle Botha, Sandra Barnard, Morné du Plessis, Mushal Allam, Kai Behn, Arshad Ismail, Anja Linstädter, Malesela Vincent Mokoka, Zamantungwa Thobeka Happiness Mnisi, Frances Siebert

Data type: png

Explanation note: Bar plots generated from the taxonomic assignments of the Welverdiend rbcL (*rbcL*_WV) and trnL dataset (*trnL*_WV) showing the absolute abundance of reads assigned to families (A and B) and genera (C and D) identified as well as the proportions of unassigned reads.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.127959.suppl3

Supplementary material 4

Bar plots generated from the taxonomic assignments of the Syferkuil rbcL (*rbcL*_SK) and trnL dataset (*trnL*_SK)

Authors: Danielle Botha, Sandra Barnard, Morné du Plessis, Mushal Allam, Kai Behn, Arshad Ismail, Anja Linstädter, Malesela Vincent Mokoka, Zamantungwa Thobeka Happiness Mnisi, Frances Siebert

Data type: png

Explanation note: Bar plots generated from the taxonomic assignments of the Syferkuil rbcL (*rbcL*_SK) and trnL dataset (*trnL*_SK) showing the absolute abundance of reads assigned to families (A and B) and genera (C and D) identified as well as the proportions of unassigned reads.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.127959.suppl4